Language In My Mouth

Linguistic Variation in the Nmbo Speech Community of Southern New Guinea

Eri Kashima

A thesis submitted for the degree of
Doctor of Philosophy
of The Australian National University

Australian National University

February 2020

This thesis represents an original piece of work, and does not contain, in part or in full, the published work of any other individual, except where acknowledged.

Eri Kashima
February 2020

Abstract

This thesis is a mixed-methods investigation into the question of the sociolinguistics of linguistic diversity in New Guinea. For many decades there have been suggestions that something about the societies and cultures of New Guinean speech communities are conducive to language differentiation and diversification (Laycock 1991, Thurston 1987, 1992, Foley 2000, Ross 2001). There have, however, been few empirical studies investigating these claims. In this thesis I investigate linguistic micro-variations within a contemporary New Guinean speech community. The goal is to identify socio-cultural selection pressures that effect language variation and change, which can be scaled up to understand the linguistic diversity we see today in New Guinea.

The community under investigation is the Nmbo speech community, located in the Morehead area of Southern New Guinea. It is a highly multilingual community in the middle of the Nambu branch dialect chain of Yam languages, and consists primarily of the three villages Govav, Bevdvn, and Arovwe. The ideologically licensed speakers of Nmbo are the Kerake tribe people, but due to the practice of marriage exogamy, a large portion of non-Kerake people speak Nmbo as an additional language learnt from their parents, or spouse. This thesis embraces the complexities of the multilingual ecology by including data from Kerake women who have married out of the Nmbo villages into the neighbouring Nen language village of Bimadbn.

The empirical investigations bring data from three directions. First, the qualitative descriptions based on my own ethnographic fieldwork supported by prior ethnographic descriptions of the area. The picture to emerge is that of an egalitarian multilingual speech community (François 2012) with multilingual speakers of indigenous vernaculars. The qualitative descriptions also provide basic facts pertaining to the demographics and social structure of the community, and the Morehead area more broadly. Second is the linguistic description of the Nmbo

language. Nmbo is an under-described language, and this thesis contains a sketch grammar covering the basic aspects of Nmbo grammar. Finally there are three quantitative studies of variation. The vowel sociophonetics study and the word initial [h]-drop study are classic Labovian variationist studies that investigate patterns of variation across a sample of speakers. The former is based on elicited word list data, and the latter on naturalistic speech data. The third quantitative study takes a grammaticalisation approach to investigating an emergent topic marker in a topicalising construction from a relative clause construction.

This is the first thesis ever produced providing qualitative, descriptive, and quantitative data from a New Guinean speech community within a language ecology of vital vernacular multilingualism. The contributions of the thesis are twofold. Firstly, this thesis brings grammatical and sociolinguistic descriptions from an under-studied language, and region of the world. It is a socio-grammar (Nagy 2009) that considers language ecology, sociolinguistics, and grammatical description. Secondly, this thesis contributes empirical data to the sociolinguistics of small-scale speech communities. The classic sociolinguistic variable of gender is not found to be particularly significant in the variables studied, despite the community being highly gendered in other social domains. Village, however, shows some significance. Overall the Nmbo speech community has relatively homogeneous ways of speaking as far as the three variables are concerned, and paint a picture of a tight-knit society of intimates (Trudgill 2011).

The conclusion to the question of the sociolinguistics of diversification in Nmbo is that while there is some evidence of sociolinguistic differentiation within the speech community, the most important social groups to differentiate against are the other sister language groups and tribes in the immediate vicinity. The nascent variation within the Nmbo speech community, combined with the ethnographic evidence of a cluster of dense and multiplex social networks, suggest that if the social need to differentiate between other Kerake should arise, linguistic differentiation may occur rapidly.

Acknowledgements

The words for this thesis may have been typed out by myself, but it is unequivocally produced as an artefact of team work.

First my thanks go to my supervisor Professor Nicholas Evans, who has given me this life-changing opportunity to work on such an inspiring and broad-perspective project under the banner of the Australian Research Council Laureate Project: The Wellsprings of Linguistic Diversity (Grant No. FL130100111). In the same vein I would like to thank the ARC Centre of Excellence for the Dynamics of Language for the additional support I have received. I have been utterly spoiled by the opportunities and encouragement received over the duration of this thesis project.

Secondly, my sincerest thank you goes to the Kerake people of the South-Fly District of Western Province, Papua New Guinea. They have been the most patient and forgiving field site a first timer could have asked for. Thank you to all the villagers of Bevdvn who housed and fed me, and who bore a disproportionate amount of labour educating me, caring for me, and looking out for me. Thank you to the villagers of Govav and Arovwe, the latter with whom I could not work with as much as I would have liked. Thank you to the Yarne tribe people of Drdr and Pongarki, and I hope the future brings opportunities for linguistic work to be conducted on such an interesting community. Thank you to all the speakers who contributed their narratives to the Nmbo Sociolinguistic Corpus, and the patient speakers who contributed to the Nmbo Word List Corpus. As I hear more about other linguists' experience in conducting word list elicitations in other parts of the world, I realise I was blessed with a patient community. Thank you also to the villagers or Bimadbn for having me over, particularly in the years 2014 and 2017.

A special thank you goes to Bunai Aniba, without whom the transcription and

translation of the Nmbo Sociolinguistic Corpus would not be in as good a shape as it is currently. A special thanks to the other key transcribers and translators: Lawrence Mike, Kawas Säme, Gima Zoga, Richard Gima, Rhouda Timoti, Jacklyn Yavs, Rachel Midawa, Towal Taga, Lazarus Smako, Tano Sam, Mwigm Nuwira, Anna Dawi, Toka Arura, Wayni Mege, Ella Awasi, Lazarus Smako, Kata Suga, Karuwa Wagra, Weko Sovam, Zoga Gima, Yumer Yäkimo. Personal thanks go to Ene Aniba, Towal Taga, Doado Suga, Bézbéz Taqm, Kaune Saka, Sosem Baoto who helped me so much in my day to day logistics of living in Bevdvn. Thank you to the Kerake women residing in Bimadbn: Fangore Yaki, Ḡño Nézén, Rucien Aniba, Gerida Nuwira, Yayam Smako, Hom Yavs, Bethsheba Zoga.

An extended thank you to my linguist colleagues at the ANU and across other institutions. My fellow Wellspringers at the ANU have been my emotional and intellectual rock: Marie France Duhamel, Alexandra Marley, Saliha Muradoğlu, Hedvig Skirgård, Catherine Travis, and Bianca Hennessey. Thank you to other geographically distant Wellspringers: Sonja Gippert, Murray Garde, Simon Greenhill, Miriam Meyerhoff, and Ruth Singer. A special thank you to Wellspringer T. M Ellison who has irrevocably altered the way I view and approach language phenomena. A very special thank you to fellow Wellspringer and Southern New Guinea fieldworker Dineke Schokkin, it has been wonderful to have someone so close who is both part of Southern New Guinea fieldwork and the Wellsprings project.

Thank you to the other Southern New Guineaists, Christian Döhler, Jeff Siegel, Julia Miller, Kate Lindsey, Kyla Quinn, and across the border to Wayan Arka, Matt Carroll, Tina Gregor. A special thank you to readers of earlier draft versions and sections: Danielle Barth, James Grama, and a big thanks to Bruno Olsson. Thank you to other ANU linguists past and present who have all given me inspiration, guidance, and feedback along the way: Cynthia Allen, Avery Andrews, Wolfgang Barth, Matthew Callahan, Claudia Cialone, Greg Dickson, Owen Edwards, Kristina Gallego, Jennifer Hendriks, Darja Hoenigman, Huade Huang, Simón Gonzáles Ochoa, Siva Kalyan, Sydney Kingstone, Evan Kidd, Naijing Liu, John Mansfield, Kate Naitoro, Lauren Read, Malcolm Ross, Lauren Sadow, Jane Simpson, Kwangju Soh, David Nash, Carmel O'Shannessy, Charlotte van Tongeren, Melody Ann Ross, Yoko Yonezawa. Thank you to colleagues across Australia and the world who also have inspired me: Rosie Billington, Josh Clothier, Anne-Laure Dotte, Wilfred Fimone, Ake Nicholas, Patty Epps, Geof-

frey Haig, Lauren Gawne (I wouldn't have applied to the Wellsprings project if it weren't for you), Janet Fletcher, Katie Jepsen, Ivan Kapitonov, Barb Kelly, Felicity Meakins, Kazuhiro Ogura, Shana Poplack, Sally Rice, Stephan Schnell, James Stanford, Peter Trudgill, Tony Woodbury. Thank you also to Jo Allen, Jo Bushby, Jessica Chen, Katie Cox, Ally Wolfe, Victoria Firth-Smith, Etsuko Mason, Tobias Maltez, Stephen Mann, Inger Mewbern, Romina Paskotic.

Thank you to the three examiners of my thesis, whose critiques, insights, and comments have made better the final version of the thesis.

Out side the world of linguistics, I must give special thanks to Vikki Ong for giving me a life outside academia. Also big thank you to Liliana Braumberger, Alex Carley-Hulyer, Sarah Komatsu, Nyx Matthews, and Jennifer Phan.

Thank you to my parents Yoshi and Emi for raising me essentially to write this thesis, and offering both academic and non-academic advice.

The final special thanks goes to Paul, for coming to the Morehead area, and for your endless patience and encouragements even though you couldn't give two hoots about vowels, verbs, or grammar.

Kerake Language Committee

Thank you to the various members of the Nmbo Zi Committee. From each village, in alphabetical order, with the years of membership:

Arovwe

Mr. Bunai Aniba (2015-2017)
Mr. Kaog Patra (2014-2017)
Mrs. Rhouda Timoti (2015-2017)
Mrs. Qai Timoti (2016-2017)
Ms. Toni Peter (2014)

Bevdvn

Mr. Baoto Dimban (2014-2017)
Ms. Jacklyn Yavs (2014-2017)
Mrs. Rachel Midawa (2014-2017)
Mr. Tano Sam (2014-2017)

Govav

Ms. Emily Daumi (2015)
Mrs. Jenny Abao (2017)
Mr. Gima Zoga (2016-2017)
Mr. Gima Mwigm (2014)
Mrs. Wrkea Rami (2017)
Mrs. Märä Wagra (2017)
Ms. Melcy Wasang (2015-2017)
Mr. Mwigm Nuwira (2014-2017)
Mr. Richard Gima (2015-2017)

Members of Supervisory Panel

Professor Nicholas Evans (Primary Supervisor and Chair of Panel)

Professor Catherine Travis (Associate Supervisor)

Dr. Timothy Mark Ellison (Associate Supervisor)

Dr. Bethwyn Evans (Associate Supervisor)

Dr. Simon Greenhill (Associate Supervisor)

Contents

Abbreviations

\.../	Verb stem
±	Plus/Minus
α	a-set verbal undergoer prefix
β	β-set verbal undergoer prefix
A	Actor
ABL	Ablative
ABL.DEM	Demonstrative Ablative
ABS	Absolutive
AG.NMZR	Agent Nominaliser
ALL	Allative
ALONE	Alone suffix
ALR	Already (iamitive)
AND	Andative
APP	Applicative
a-suffix	Actor suffix
BEN	Benefactive
COM	Comitative
CONT	Continuative
DAT	Dative
DEM.dist	Distal Demonstrative
DEM.fw	Forward manner demonstrative
DEM.mn	Manner demonstrative
DEM.prox	Proximal demonstrative

DU	Dual
DTRN	Detransitiviser
EMPH	Emphatic
ERG	Ergative
ERG.nsg	Ergative Non-singular
EXPL	Expletive
FRUS	Frustrative
FUT	Future
GOAL	Goal
HAB	Habitual
HYP	Hypothetical
IMP	Imperative
INTS	Intensifier
IPFV	Imperfective
LOC	Locative
LVC	Language variation and change
NDU	Non-Dual
NEG	Negation
NPHD	Non-Prehodiernal TAM
NSG	Non-singular
ONLY	Only suffix
ORI	Originative
P	Patient
PFV	Perfective
PHAB	Past Habitual
PL	Plural
PL+	Large plural
POSS	Possessive
POSSC	Close Possessive
PRET	Perfective Preterite

PRIM	Primordial TAM
PRIV	Privative
PROXIN	Proximate Inceptive TAM
PST	Perfective Past TAM
PURP	Purposive
QUANT	Quantifier
QUOT	Quotative
RDP	Reduplication
RMPST	Remote Past TAM
RST	Restrictive
S	Sole argument
SG	Singular
SIM	Similative
SUP	Supposition
TEMP	Temporal marker
TEMPG	Temporal goal marker
TOPIC	Topic Marker
U	Undergoer
USM	Utterance Selection Model of Language Change
u-prefix	undergoer prefix
VEN	Venitive
YPST	Yesterday Past TAM

List of Figures

List of Tables

Notes on orthography

Nmbo is one of the few languages in the Morehead area which had an orthography developed prior to 2008. It was developed by a linguist or a team of linguists at the Summer Institute of Linguistics, whom I have not been able to trace down at this point in time. This section makes reference to the orthographic conventions laid out in a dissertation by Martin (2001), though he does not give the reference of who developed the orthography.

One of the particularities of Nmbo and its sister languages is the presence of short vowels which are mostly predictable and epenthetic (section 4.1.2). The short vowels function to break up illicit consonant clusters in onset and coda positions (list of licit clusters are presented in the appendix A.1). /ɐ/ is the more frequently occurring of the two short vowels, and will be represented orthographically by an absence of a symbol in a word:

e.g. *kanam* (/ka.nam/ 'snake') vs. *knm* (/kɐ.nɐm/ imperative 'come')

The other reduced vowel, represented in IPA as /ə/, occurs less frequently. It does, however, form minimal pairs, and so will be represented orthographically by the symbol < é>.

e.g. *yam* (/jam/, 'custom', 'way', 'event') vs. *ym* (/jɐm/ , 3sg.NPST/be) vs. *yém* (/jəm/, 'gumtree').

The following table summarises the three-way correspondence of IPA symbols to SIL Orthography and to the New Orthography used by Evans in the documentation of Nen, the neighbouring Nambu branch language to the east. I have taken some liberties with the interpretation of the SIL vowel orthography correspondences, since the SIL analysis of the vowel space appears to differ from mine. Currently the New Orthography is being introduced into the Kerake villages, and as far as I can tell, both community and evangelical missionary are happy to

adopt this style. Aniba and Martin have recently produced a few chapters from the Gospel of John in the New Orthography, and a few Nmbo and Nen speakers are messaging via the internet in a variation of the New Orthography.

IPA	k͡p	g͡b	ɲ	ɸ	β	j	i	e	æ	ɑ	o	u	ə	ɐ
SIL Orthography	kw	gw	ny	f	v	y	i	e	ee	aa	oo	u	o	a
New Orthography	q	ḡ	ñ	f	v	y	i	e	ä	a	o	u	é	á*

Table 2: IPA to orthographic correspondence for non-latin characters.
*Used only when word initial, or finally in a monosyllabic open syllable word.

Examples of New Orthography in relation to IPA symbols:

(1) a. tande ytqn Ḡño ym.

tan.de jɐt.k͡pɐn g͡bɐ.ɲo jɐm.

My name is Ḡño .

b. wén tomba tavotan zäkot.

wən tom.ba ta.βo.tan zæ.kot.

I chopped down a long tree to use as a jack.

c. tngvrendawn ärägv.

tɐn.gɐβ.ren.da.wɐn æ.ræ.gɐβ

I crossed over the creek.

Notes on reading the interlinear glossing

The convention of Yam language verbal interlinear glossing is designed to take account for the circumfixal nature of the verbal morphology, in addition to the distributed nature of argument and TAM information across multiple affixes (section 4.5).

The notation ‘ \/ ’ marks the stem of the verb, with root in-between the slants, and the argument + TAM affixes on either end of the root. In the examples below I have **bolded** the verb root in the orthographic representation, and in the interlinear gloss row.

The interlinear gloss for the circumfixal notation gathers up the argument information in terms of person and number, and semantic role. In example 2a with a divalent (transitive) verb, the ‘ 1sg.A>3pl.U ’ is to be read as ‘first person singular actor is acting upon a third plural undergoer’. In example 2b with a monovalent (intransitive) verb, only the sole argument is expressed in the interlinear gloss, i.e. 1sg.A.

The TAM information is gathered after the argument information. Section 4.7 presents a quick overview of the abbreviations and semantics of TAM. Directional information is also marked after TAM information as VEN (venitive ‘hither’) and AND (andative ‘away’), if present (example 2b).

(2) a. yn**mi**/tan

1sg.A>3pl.U:ipfv.nphd/**ask**

I am asking them.

b. n-ng**ango**/tan

1sg.A:ipfv.nphd:and/**return**

I am returning.

The more conventional morpheme-by-morpheme glossing will be used in sections where it is pertinent to the point. The symbols 'α' and 'β' in the undergoer prefix slot indicates the prefix set to which that prefix belongs (section 4.5.3). In example 4 the 'M' of the undergoer prefix indicates a person/number invariant middle prefix, meaning that the predicate is monovalent.

(3) a. n-ng-**ango**-ta-n

M:α-and-**return**-ipfv.ndu-1sg.A

I am returning.

b. k-ng-**ango**-ta-n

M:β-and-**return**-ipfv.ndu-1sg.A

I was returning yesterday.

Nmbo speakers often omit the pronominal form of a referent, or will use anaphoric mechanisms to refer to various entities and occurrences. In the English translations of examples, I will often include contextual information in square brackets '[]' to aid interpretation. In some examples, I will also give a literal translation in the square brackets which better reflect the original Nmbo, where the English translation gives a better approximation of the sense in English.

(4) a. n-ng-**ango**-ta-n

M:α-and-**return**-ipfv.ndu-1sg.A

I am returning **[from the garden]**.

b. kruvr-am de w**ivo**/ø

cold-ERG ALR **3sg.A**>1sg.U:pfv.pst/finish

I was cold. **[lit. The cold finished me off.]**

(WSEK1-G20170723-01GwN10KK, 00:02:30.412 - 00:02:31.247)

Notes on data source

Examples used in this thesis contain information about the source of the data. A code beginning with *WSEK1* indicates the PARADESIC (2019) digital archive collection ID. Notes from my field notes will be indicated with a preamble "Fieldnotes" and year of field work. At the time of this thesis submission, the Nmbo Sociolinguistic Corpus will pending submission and processing with PARADESIC.

Information on the source of translation is indicated for the qualitative sections of the thesis. The initials of the translators are provided, and match the names of the Kerake Language Committee members listed in the acknowledgements. The initials EK refer to myself.

Details of the composition of the Nmbo Wordlist Corpus are presented in the appendix (A.16), and likewise for the Sociolinguistic Corpus (A.6, A.7).

Chapter 1

Introduction: The Question and Approach

> "Melanesia shows the world's greatest linguistic diversity... A cardinal problem in the study of the prehistory of Melanesia is the formulating of some hypothesis as to the reason for this diversity."
>
> *Laycock 1991:31*

Laycock begins his paper with a contemplative tone, belying the conviction of his argument in the rest of the article. He argues that New Guinea has high levels of linguistic diversity because there are strong socio-cultural pressures towards language divergence, and the maintenance of sociolinguistic differences. Many who have worked in the region make arguments similar to Laycock (Thurston 1987, 1992, Foley 2000, Ross 2001), that something about New Guinean communities and their cultures play a part in language diversification. To date, however, there have been few empirical studies at the speech community level which investigate how socio-cultural variables relate to diversification through the process of cumulative language variation and change (Evans 2018:26-31).

This then raises the question of how one should empirically study these socio-cultural parameters that affect language change, and language evolution. This thesis project is one such attempt, which investigates the sociolinguistics of lan-

guage variation and change in the speech community of Nmbo[1] [nɐm.bo] speakers of the Morehead area, near the southern coast of Papua New Guinea. It is a mixed-methods investigation within an overarching cultural evolutionary framework. I present qualitative, quantitative, and descriptive linguistic data in order to identify and contextualise socio-cultural variables that affect language variation and change. This thesis aims to answer the question:

> What socio-cultural pressures are affecting language variation and change in a New Guinean speech community today, which can be used to understand historical language evolution in this linguistically diverse part of the world?

The choice of Nmbo as a speech community was not accidental. Not only is the Morehead area of Southern New Guinea a highly linguistically diverse part of New Guinea, it is also a community that retains a traditional language ecology of indigenous multilingualism. Nmbo is a vital language with some 700-1000 speakers living primarily in the three villages of Govav, Bevdvn, and Arovwe, with many Nmbo speaking women living in adjacent non-Nmbo villages due to the convention of virilocal exogamy. It is one of the few remaining places in the world where multilingual practices are not heavily influenced by the presence of an incoming world or national language, and revolves around the vernacular languages which are all spoken in vital but small communities. For the Kerake[2] tribe people, the autonym of the peoples who claim ownership over the Nmbo language, this language ecology is a natural part of their life. The following words by Kata, a man in his late thirties from Bevdvn, captures in brief how communication is undertaken in the multilingual landscape of the Morehead area. His words also express the intertwined concepts of language and place, which are intrinsic to his sense of self. The title of this thesis was inspired by his turn of phrase, *brmbran tvende zi km bä bädi ymn*, "our language-voice (melody) is in the mouth":

> *Bimadvnat ge kngovarn ämb evhon ädi nowavtam ymovayo, yvende zie. Nen zie nngowavtam. Ämb taiman giymn, yao kor nngowavtan däv. Tvende zi gs ymm, brmbran tvende zi km bä bädi ymn, deutut yao bä kevonga. Mnan ynd bä yna anohmn är wm, tande zian, Nmbo*

[1] Glottocode NAMB1293, ISO-639-3 code *ncm*.

[2] Also known as Keraki in the classic ethnography of Williams (1936). I have chosen the orthographic representation *Kerake* to better approximate the pronunciations of Nmbo speakers I heard in the field.

zian nanotamn.

> "Sometimes when I arrive to Bimadvn [the village of Nen speakers], we all speak together with them, using their language. We speak the Nen language, other times I won't really speak it. Our language, ok, it is really part of us [lit. our language-voice (melody) is in the mouth], it is forever and will not end. Because I am a person grown of this place, I have grown with my Nmbo language."

Translation LM, KS2, major edits by EK. WSEK1-B20150908-04PerceptualKS, 00:06:37.151 - 00:06:54.533[3]

The existence of multiple languages in the Morehead area is also a source of deep interest for the Kerake. A typical contemporary answer to the question "why are there so many languages in the Morehead area?" can be seen in the following words by Hom, an elderly woman who married out of her Nmbo village into her husband's village decades ago:

> *Yä wavroärum ädi yna zi tngmendao. Dena de. Yä nmbo zi, yä namna zi [...] Gänzänam yna zi tngmendao. Yao tvende kakivem. Yä gänzänam tmendao.*

> "He the creator man, he laid down these many languages. Like this. He gave the Nmbo language, he gave the Namna language [...] God put down these many languages. Not our ancestors. God put them down."

Translation by JY, RM, and EK, WSEK1-B20150813-03PerceptualHZ

This passage reveals interesting aspects about Nmbo, both socio-cultural and linguistic. The translations of the lexical items *wavroär* as 'creator man', and *gänzän* as 'god' were given by two women in their early thirties who have grown up in a community that is nominally Christian. Like many parts of New Guinea, however, the interpretation of Christian stories and doctrine are understood through the lens of pre-existing cultural and narrative types. *Gänzän* is a word also used when referring to the primordial spirits of traditional Kerake myths; one of whom was responsible for populating the Morehead area with speakers of many languages. Missionary work has been sporadic in the Morehead area, and while it has left a mark on the cultural landscape, the impact is still shallow

[3]The information provided within in-text examples will include the source from which the example was taken, and occasionally, information on who provided the translation. More information about data sources is presented in the preamble of the thesis.

time-wise. Elderly Kerake still recall non-Christian traditions, and were keen to re-enact and document these stories and practices. For the most part the Kerake lead a way of life which was shallowly affected by colonisation.

There are also some linguistically interesting features in this passage. The verb *tngmdetao* [tɐn.gɐm.de.tao] marks the undergoer partially by the leftmost prefix *t-*, and the actor partially by the rightmost suffix *-tao*. When viewed together, both affixes contribute TAM and argument information. In other words, the two affixes form a circumfixal paradigm of core argument marking and TAM. This verb also contains an andative 'away' prefix *ng-* which is sometimes used to mark large plural number (noted in the sister language Nen, Evans 2019c). Nmbo is a language with rich verbal morphology, exhibiting cross-linguistically unusual features like the use of the andative for the expression of a large plural.

In addition to the interesting language ecology, Nmbo is an under-described Papuan language without substantial prior description.[4] Work is increasingly being conducted in the area of Southern New Guinea (e.g. Evans, Arka, et al. 2018), with substantial work done on Nambu branch languages of Nen (Evans 2012a, 2015a, 2015b, 2016, 2017, 2019) and Nama (Siegel 2014b, 2016, 2017). An equally important goal of this thesis is as a proof of concept of doing language documentation concurrently to documenting sociolinguistics (cf. Childs, Good, Mitchell 2014), as well as studying variation (Hildebrandt, Jany, and Silva 2017); what Meyerhoff (2017) poetically describes as "writing a linguistic symphony" as opposed to a "sonata of individual variables". The linguistic description and studies of variation in this thesis are based on eight months of fieldwork across the years 2014-2017, with trips ranging from two weeks to 3 months. While the main research goal of this thesis is to contribute to the big-picture question of language diversity, the related secondary goal can be expressed as the creation of a descriptive socio-grammar (to borrow a term from Nagy 2009). An account of language change necessarily involves understanding both socio-cultural and linguistic factors, which this thesis aims to describe.

This introductory chapter will outline the rationale behind conducting a synchronic study of variation and change for the purposes of understanding the diachronic phenomena of language evolution and linguistic diversity (section 1.1). I present the use of a cultural evolutionary framework as a way of linking macro-

[4]Some non-linguistic analysis of temporal semantics has been undertaken by Martin (2001), and some lexical similarity comparisons have been made by Rueck (2006), and Tucker, Boevé, Fuller, Gustafsson, and Rueck (2003)

and micro-variation, which also unifies the various methodologies employed in linguistics to study variation and change ranging between variationism and grammaticalisation (section 1.2). I will then review existing hypotheses pertaining to the role of socio-cultural pressures in language change, and highlight the specific socio-cultural pressures that have been proposed as relevant for New Guinea (section 1.3). The methodology section will briefly outline the methods employed in this project, and introduce the linguistic variables under investigation (section 1.4). This chapter will conclude with an outline of the structure of the thesis (section 1.5).

Before continuing, a couple of notes about terminology are in order.

1) The term *New Guinea* will be used to refer to the combined geographic area of the independent state of Papua New Guinea and the Indonesian side of the island. New Guinean languages will therefore be used to mean the languages that are spoken across this geographic spread, so will include Austronesian and non-Austronesian languages. The term *Papuan languages* will be used when making specific reference to the non-Austronesian languages, but with the understanding that these do not form a single phylogenetic unit (Foley 1986, Ross 2005). *Papua New Guinea* or *PNG* will be used to refer explicitly to the territories of the independent nation state of Papua New Guinea.

2) When a language is introduced in text, the name and classificatory information will be given. The parenthesis following the language name will contain phylogenetic information with the highest level and lowest known level. Geographic information will be provided after the phylogenetic information, separated by a semi-colon.

e.g. Australian English (Indo-European, West Germanic; Australia).

1.1 Rationale: The relationship between linguistic diversity and language change

There are multiple measures for what is called *linguistic diversity*, but New Guinea is often crowned *the* area with highest diversity. Nettle (1999) suggests three types of linguistic diversity (pp.10-11), of which New Guinea shows high levels for all three. *Language diversity* refers to the number of identified languages in a region. New Guinea has a high level of language diversity with between 850

to 1300 languages (Nettle 1999:10 for the former estimate, Palmer 2018:1 for the latter). According to Nettle (1999), if the language-to-population ratio of New Guinea is applied to the United States, there would be 60,000 independent languages across the fifty states (p.3). The Amazon and Orinoco basins come second in terms of language diversity (some 300 languages and 50 families, Rodrigues 2000 cited in Epps and Salanova 2013:1), but as Dahl (2008) notes, these languages are spread across a geographic area much wider than that of New Guinea (pp.217-8). *Phylogenetic diversity* is the number of language lineages found in an area. The most recent comprehensive introduction of Papuan languages states there are 43 distinct families and 37 isolates (Palmer 2018:6). The combined number of 70 language families and isolates exceeds the estimated 50 distinct families posited for the Amazon and Orinoco basin mentioned before. These figures show that New Guinea has levels of phylogenetic diversity unmatched in other parts of the world with comparable geographic sizes. Finally *structural diversity* (or what Evans 2018:11 calls "structural disparity") refers to the variety of linguistic structures found in a region. The structural diversity of New Guinea is thought to be very high with a vast range of phonological and morphosyntactic phenomena across language families (Foley 2000). Foley (2017) gives an example of morphological disparity with Yimas (Lower Sepik-Ramu) described as having a "polysynthetic profile", while Kimaghama (Kolopom, Trans-New Guinea) is isolating (p.895). An example of phonological disparity can be seen in comparing Yélî Dnye (Papuan, Isolate; West New Britain) with its phonemic inventory of 38 vowels (Henderson 1995:3) and 56 consonants (p.10), while Aita Rotokas (Papuan, North Bougainville; Bougainville) has a total phonemic inventory of fourteen: five vowels and nine consonants (Robinson 2006). These examples suggest a wide range of linguistic features are to be found in New Guinea and within Papuan languages.

The explanations offered for New Guinean linguistic diversity come from an evolutionary perspective. The contemporary distribution of languages is a result of cumulative historical changes, so in order to understand synchronic diversity we must understand how languages changed in the past. Language change is increasingly being discussed in terms of a process of cultural evolution (Croft 2000, 2006, 2008, Mufwene 2001, Atkinson and Gray 2005, Evans 2013, Greenhill 2015). While languages and other cultural constructs are not the same as biological entities (Croft 2000, Dediu et al. 2013, Evans 2019a), it has been recognised that language and cultural evolution follow a generalised model of evolution.

That is, biological evolution and cultural evolution follow broadly similar patterns of selection and reproduction, thought the precise means by which selection and reproduction occur are specific to the system in question (the parameters particular to language evolution will be outlined further in section 1.2). Framed this way, the question of linguistic diversity can be tackled as a question of how evolutionary parameters differ across the world. For example, high levels of linguistic diversity in an area can be understood as "the outcome of differences in the birth-rate and the death-rate of languages" (Greenhill 2014:557). New Guinean linguistic diversity can therefore be hypothesised to result from high rates of language change and language birth. The more variation and change is inherent in a population, the more opportunities there are for speciation (Gillespie 1998:5). Evolutionary biologists use the term *tempo* when referring to these rates of change. Tempo, however, is affected by a variety of parameters, which evolutionary biologists call *mode*. Greenhill, citing Simpson (1944), describes mode as "how populations became... differentiated, ...how they passed from one way of living to another, or failed to do so." (2014:558).

A variety of linguistic modes have been proposed for increased tempo. The type of linguistic structure (i.e. phoneme, phrase structure, lexical item), and the different subsystems of a linguistic system (i.e. phonology, morphosyntax, semantics), are known to change at different rates (Nichols 1992, 1995, Dediu and Cysouw 2013). For example, Wichmann and Holman (2009) found that pragmatic features related to discourse organisation tend to be unstable compared to morphosyntactic features. Dediu and Cysouw (2013), comparing the findings of five different studies, report that some features were recurrently found to be stable,[5] but the unstable features spanned the entire gamut of linguistic subsystems. A Bayesian phylogenetic test of 81 Austronesian languages suggested that basic vocabulary is more stable over time than various grammatical structures, i.e. that grammar changes faster (Greenhill et al. 2017).[6] But many of the findings based on global language samples are difficult to interpret if we try and generalise across regions and linguistic areas.

[5]The three mentioned in-text are: order of adjective and noun, order of genitive and noun, absence of common consonants (Dediu and Cysouw 2013:16)

[6]There are, however, also studies that raise questions about the veracity of the stability of basic vocabulary (Tadmor, Haspelmath, and Taylor 2010). Wichmann and Holman (2009) suggest that while some basic vocabulary items in a Swadesh list are stable, typological features such as those in WALS may be better suited for inferring phylogenetic relationships between languages.

Further more, we do not know whether the linguistic features of New Guinean languages change faster than languages of other parts of the world. Comrie and Cysouw (2012) identify a list of "typical characteristics of Papuan languages" (p.85) when viewed through the feature lists of WALS, the World Atlas of Language Structures (Haspelmath, Dryer, Gil, and Comrie 2008), but they conclude these are not so cross-linguistically unusual (p.88). The WALS feature list excludes typological rarities by design, however, so we cannot tell from Comrie and Cysouw's study whether New Guinean languages have features that are more-or-less susceptible to rapid rates of change. Our understanding of the structural aspects of New Guinean languages is low compared to other parts of the world due to the low coverage of descriptions (Hammarström and Nordhoff 2012, Evans and Klamer 2012), but this situation is rapidly improving in recent years (e.g. The regional sketches in Palmer 2017, reference grammars by Döhler 2018 for Komnzo, Sarvasy 2017 for Nungon, Olsson 2017 for Costal Marind, Wilson 2017 for Yeri, Pennington 2016 for Ma Manda, Honeyman 2016 for Momu). The high levels of structural disparity that we already know about, however, suggest that linguistic factors alone can only explain so much in terms of tempo of change. There is increasing recognition that effects of areality, linguistic genealogy, and chance happenings, need to be recognised as a part of understanding global language distributions (Roberts and Winters 2013, Greenhill 2014, Hickey 2017, Hua, Greenhill, Cardillo, Schneemann, and Bromham 2019).

Other suggestions pertaining to tempo concern historico-demographic factors, such as settlement times of people. Places where speech communities have settled for a long time have been hypothesised as having more languages. Nichols (1992) uses the term *language residual zone* to describe, among other things, areas of high linguistic diversity where languages have accreted over deep time, resulting in an eventual net gain of languages in the area. Archaeological evidence suggests settlements of the New Guinean highlands occurred around 50,000 years ago (Summerhayes et al. 2010, Allen and O'Connell 2014), supporting the scenario of long-term language accretion. But the southern coast of New Guinea is estimated to have formed as a landmass only 10,000 years ago (Hitchcock 2010:73), which in the scheme of things it not very long ago. Australia and Africa have evidence of human settlements for comparable or longer spans of time, and while these areas are also quite diverse, they do not reach the levels of New Guinea. Australia is thought to have been settled by at least 65,000 years ago (Clarkson et al. 2017), while Africa was the homeland of Homo sapiens, with molecular dating suggesting

habitation from about 195,000 years ago (McDougall, Brown, and Fleagle 2005). Time depth can explain how diversity may result by language accretion but it does not tell us how languages diversify in the first instance.

More recent investigations into demographic modes suggest that the role of community configuration and social interaction affect tempo. Linguistic features of small languages are often quite different to those of large languages, and there have been increasing attempts to identify what social parameters might be relevant in explaining these linguistic differences (e.g. Perkins 1992:92, Kuster 2003:41-5, Sinnemäki 2009, Bowern 2010). Trudgill (2011) has identified five characteristics of what he calls *societies of intimates* for this end. Trudgill's interests lie in understanding language complexity, but his list of socio-cultural parameters are applicable for schematising language change more broadly. Trudgill hypothesises that societies with the following five social features are conducive to language complexification: 1) low levels of adult language contact, 2) high social stability, 3) small population size, 4) dense social networks, and 5) large amounts of community-shared information (2011:147). Of these features, population size and social interactions (i.e. social networks and community-shared information) within small groups are thought to be particularly relevant in understanding language diversification. Dale and Lupyan (2012) note that small languages are often more inflectionally complex, and many scholars are in support of the hypothesis that it is because small populations allow learners of a language to have iterated face-to-face interactions over the course of their life time to master these complexities (Reali, Chater, and Christiansen 2014, Lupyan and Dale 2010, Wray and Grace 2007). Work from experimental semiotics also suggests that the demands of interaction shape the form of symbolic systems, with evidence that group size affects the type of interactions people have within a group (c.f. Michael 2014:489-491). For example Fay, Garrod, and Carletta (2000) suggest that smaller groups have more dialogue-like interactions, while larger groups have more monologue-like forms of communication. Over time these interaction styles are thought to have consequences for the evolution of the communicative system.

These investigations into the modes of community configuration and social interaction patterns are an important step along the way to link macro-level linguistic diversity to the meso/micro levels of speaker-based models of language change. Indeed, the suggestions offered by fieldworkers such as Laycock (1991) and Thurston (1987, 1992) regarding New Guinean diversity pertain to charac-

teristics of the speech communities. The various subfields of linguistics tackle language change in their own way, but I will suggest that taking an evolutionary approach, such as on a generalised model of evolution as articulated by Croft (2000). Croft's model is a useful way of bringing together common findings and approaches that hypothesise socio-cultural factors as playing a major role in long-term macro-evolution.

1.2 Frame: A Micro-Evolutionary Perspective of Variation and the Theory of Utterance Selection

Language change is increasingly viewed as a phenomenon that follows an evolutionary trajectory (Dediu et al. 2013, Greenhill 2015, Enfield 2014, Croft 2000, 2008, Currie and Mace 2012), although the degree to which selection and drift characterise language evolution is still being discussed (i.e. Newberry, Ahern, Clark, and Plotkin 2017). A widely recognised challenge in developing conceptual models of evolutionary processes is the demarcation of what constitute the units involved in the process.

I use terminology from the *Utterance Selection Model of language change* (USM) (Croft 2000, 2006, 2008) in this thesis for a few reasons. Firstly, the theory recognises that language change follows patterns of a generalised model of evolution; the concepts and terminology in this model are consistent with broader studies of evolution outside the field of linguistics. Secondly, studying evolutionary processes at the micro-level of replication links these processes to macro-level evolutionary studies, establishing a continuity between micro- and macro-change. Thirdly, much like how a macro-evolutionary perspective identifies tempo and mode as broad parameters involved in change, a micro-evolutionary model identifies and names relevant parameters of evolutionary change at a more fine-grained level of selection and replication. This helps bring together relevant findings and observations made across sub-fields of linguistics ranging from linguistic anthropology, variationist sociolinguistics, grammaticalisation studies, and many others.

Replication and selection are crucial steps that occur in evolutionary systems, and need to be addressed explicitly to understand the iterated and long-term process of change over time. Croft bases his model on Hull's 1988 *general model*

of evolution, which proposes that certain systems in the world undergo change by an evolutionary processes; biology, culture, and language being some examples of system that do. Evolutionary systems are definable and identifiable by key parameters, and it is these parameters that propel change through time. What these parameters are is dependent on the specific systems.

The goal for understanding language evolution, then, is to identify the parameters involved in a linguistic evolutionary system. Hull proposes that a general model of evolution is comprised of *interactors*, *replicators*, *population*, and *selection and selection pressures*. The *replicator* is the units which undergoes replication. In biology these are usually seen as genes and/or molecules. In language evolution, the replicator is any linguistic utterance at any level of the linguistic subsystem; an entity Croft calls a *lingueme*. The *interactor* is the entity that interacts as a cohesive whole with its environment in such a way that this interaction causes replication to be differential (Croft 2000:13, Hull 1988:408). In the case of language this is an individual speaker, a dyad, or any other conversational unit within which utterances are produced. The *population*, in a simplistic sense, is a collection of interactors who can cause replication of replicators. The equivalent of a population in Croft's view are communities that are tied in communicative interaction (2000:30); what is more commonly referred to as a speech community. Successful communicative interaction depends to some extent on the degrees of structural similarity of the varieties spoken in an interaction, but also on the social behaviour of the speakers which is often understood in a sociocultural context. The final parameter is *selection and selection pressures*. Hull defines selection as "a process in which the differential extinction and proliferation of interactors cause the differential perpetuation of the relevant replicators" (1988:409); in other words, selection is the process by which a replicators is chosen for replication over a competing one. Selection pressures are biases that exert pressure on the selection process, or more benignly, nudge selection in a certain direction.

The USM posits the utterance as the replicator. The term *utterance* is left deliberately flexible in order to capture different levels of linguistic structure; anything that can be produced in an utterance is a replicator. At the smallest level this unit may be a single phone, articulated and perceived by an individual speaker. At the largest, it may be an entire clause, or utterance made under a single intonation unit. Crucially for an evolutionary model, replicators are in competition with other replicators, and the selection of one form over another

is what cumulatively ends up forming an evolutionary system. In other words, utterances often have variable ways of being expressed. This is very similar to what Labovian language variation and change (LVC) studies call "variants of the same structure" (Labov 1965:31) or "a variant of a linguistic variable" (Tagliamonte 2012:3); some unit of linguistic structure that shows variability in its realisation. LVC studies have tended to investigate phonological replicators (e.g. /t/ vs /d/, vowel realisations), but morphosyntactic variables such as negative concord also have a history in the field (Wolfram 1969, Labov 1972a, Cheshire 1982). Historical linguistics and grammaticalisation studies tend to use a pan-linguistic term "linguistic structure" when talking about replicators, and these span the range of linguistic subsystems and structures.

The *speaker* is the interactor of a linguistic evolutionary system who produces the utterance. The speaker has at their disposal numerous options for any given utterance (i.e. variants of a linguistic variable). Since natural languages are produced linearly, only one replicator can be selected for any given utterance. The speaker replicates a replicator when they select and successfully utter it in interaction with a member of the population, the broader *speech community*. USM does not exclude the possibility that replication occurs at the level of the speech community, but treats replication as primarily occurring at the locus of individuals. The reason being that one cannot have language change at the level of community (i.e. a collection of individuals) without change at the level of individuals (Croft 2000:38). This articulation of the individual at the center of change, at least in sociolinguistics, is attributed by Weinreich, Labov, and Herzog (1968) to Paul (1880) (p.98). Additionally, the issue of what constitutes a population is in fact quite problematic in evolutionary biology and other related fields (Sober 1980, Eshel 1996), but I follow the approach outlined in USM since it does not affect the outcome or interpretation of the studies in this thesis.

Given this, the process of linguistic micro-evolution becomes a matter of how utterances are selected and replicated over time. What affects a speaker to select a certain utterance over another depends on whatever selection pressures are in operation. Evans (2013) groups a large number of selection pressures into four broad types: psycho-physiological, and socio-cultural, genetic-epidemiological, and system selection; these latter two selection pressures are less relevant to this exposition of a model of micro-variation so will be explained in a footnote.[7]

[7]Genetic-epidemiological pressures refer to the possible effect of biological tendencies in a population in the selection of certain replicators (i.e. genetic and/or prevalence of certain phys-

Psycho-physiological selection pressures typically refer to cognitive biases in cognition/processing of various types, processes involved in acquisition, and human physiology pertaining to production and processing of utterances. These include processing time required for comprehension and production of linguistic output (e.g. overlap of speech in turn-taking) (Levinson & Torreira, 2015), "chunk and pass processing" (Christiansen & Chater, 2016), lexical access and language production monitoring (Levelt 1989, Dell and O'Seaghdha 1992, Grosjean 2001, Ellison and Miceli 2017).

Socio-cultural selection pressures refer to some phenomena that arise due to "the social setting of the culture in which a language is spoken" (Evans 2013:264). Sociolinguistics as a field is defined by this core tenet of the social being at the centre of what we call language.[8] Socio-cultural selection pressures are also central to the model of utterance selection. According to Croft (2000) the identification of a new language is akin to the identification of a new species, and the criterion of language speciation is when interaction between individuals of a speech community cease; because it is then that you have two separate speech communities (pp.30-31). In a model that has speaker interaction at the heart of the replication process, and where the concept of speciation (i.e. divergence) is recognised as a separation of speech communities, socio-cultural selection pressures are claimed as the important driving force in explaining how linguistic diversification occurs (Croft 2000:31).

Here it is worth articulating what is meant by *social* and *cultural* selection pressures. Social pressures concern the behaviours of humans interacting with others individually or within a group. These include what Boyd and Richerson (1985) call "indirect bias", such as the motivation for a speaker to select certain variants over others in order to make a statement about her social identity. More generally, it can refer to speakers positioning themselves in relation to their interlocutor or other relevant social entities. Social selection pressures do not operate in a vacuum, however, and it is culture that directs individuals to make certain social decisions. Richerson and Boyd (2005) define culture as "information capa-

iological conditions in a population may explain the reason for certain linguistic typological characteristics in that population's speech). System selection refers to notions such as "system architecture", or "design features" (Hockett 1960), indicating that there are broad characteristics of any natural linguistic system that tend to preference the selection of some replicator over another.

[8]Sociolinguists have long pointed out, however, that the division between 'linguistics' and 'sociolinguistics' is disingenuous in the sense that language phenomena stems from humans and social interaction (Labov 1972b:3, Sankoff 1980c:xxi)

ble of affecting individuals' behaviour" that they acquire from other members of their species through teaching, imitation, and other forms of social transmission (p.5). Everyday words such as "idea, knowledge, belief, value, skill, and attitude" are used to describe this kind of information (Richerson and Boyd 2005:5). An individual may choose a certain linguistic utterance to make a statement about his or her social identity, but it is culture that leads them to decide that this is appropriate or beneficial in the given context.

We can apply the two concepts separately to Thurston's theory of esoterogeny (1987, 1992) by way of exposition. Based on his fieldwork in New Britain Island, Thurston (1992) proposes that languages can adapt to suit communication with community outsiders (*exoterogeny*), and community insiders (*esoterogeny*) in contact situations. If, for example, Austronesian-speaking traders arrive in a Papuan language speaking area, social pressure would bias locals to communicate with the Austronesians so they may interact with the traders for a mutually beneficial outcome. Assuming all things are equal, some communities will eventually adopt elements of the trade language as their linguistic emblem (Thurston 1987:38, 96), while other groups would not. This adoption is culture affecting the behaviours of community members. The communities who orient away from adopting useful languages such as a trade language are inclined to restrict access to their emblematic language, and such languages will become complex over time by various ways. Speakers may intentionally make their emblematic language difficult for outsiders (Thurston 1987:36), or the complexities of the language will be successfully transmitted to the next generation seeing as children are able to acquire complex languages (Thurston 1987:36). Indeed, Thurston argues that these cultural pressures explain why Anem (Papuan, Isolate) and Amara (Austronesian, Whiteman) show certain phonological and morphological complexities compared to Lusi, Kove, and Kabana (Austronesian, Siasi) spoken on the same island. The historical communities of the three Austronesian languages reflect exoteric interactions, where community members were culturally inclined to interact with outsiders. The Anem and Amara communities, on the other hand, are suggested as not so inclined.

So far in this chapter I have laid out the parameters that make up a model of micro-evolution that, in theory, scales up to long-term macro-evolution. An utterance-based theory of selection sees the speaker within an interactive speech community as the locus of replication and replicator selection. It is this speaker who is affected by a variety of psycho-social and socio-cultural selection pressures

when producing an utterance. In the next section I will delve deeper into works investigating socio-cultural pressures and language change, and discuss these studies in light of the USM with an emphasis on the socio-cultural selection pressures proposed as particularly relevant to New Guinean speech communities.

1.3 Prior Works: Hypotheses of Relevant Socio-Cultural Pressures in New Guinean Communities and other Indigenous/Minority Contexts

Socio-cultural selection pressures have often been proposed as possible drivers that affect the rate of language change in New Guinean languages. Selection pressures such as psycho-physiological ones are assumed to be constant across the human species (the *uniformitarian principle* articulated by Labov 1972c, see also Wray and Grace 2007:553), so any explanation of difference in rates of change are viewed as reflecting socio-cultural differences between communities. What we know about socio-cultural pressures on language change, however, come from a limited set of societies. As a growing number of researchers have pointed out (Nagy and Meyerhoff 2008, Stanford and Preston 2009, Childs, Good, and Mitchell 2014) sociolinguistic research has been and continues to be conducted in quite specific social settings. Researching speech communities with different social features is argued as a high priority as they will yield different kinds of sociolinguistic knowledge (Dorian 2010). This argument has been made more widely for the social sciences by Heinrich, Heine and Norenzayan (2010) who coined the useful acronym WEIRD to characterise the socio-cultural features of well-studied societies; that is, well studied societies are Western, Educated, Industrialised, Rich, and Democratic.

Research into non-WEIRD speech communities are bringing forward new data and more nuanced discussions about the role of socio-cultural pressures in language variation and change. Work on Sui speakers of rural south-western China (Stanford 2009b, 2010) shows that clan and strong identification with a paternal dialect prevents in-marrying women from conforming to the dialects of their husbands. Through her work on rural fishing communities in Scotland, Dorian (1994, 2010) found that socially homogeneous communities with closed, dense, multiplex social networks tend not to have socially-weighted variation. Satyanath and

Laskar (2009) found that historical community bilingualism between non-world languages (Indo-Aryan Bishnupriya and Tibeto-Burman Meitei) have synchronic effects in variable constituent ordering in the noun phrase. These findings support the notion that social features like cultural identity and social organisation play a role, but that these have locally particular manifestations. There are good reasons to hypothesise that socio-cultural pressures in their Papuan manifestations are central in explaining language change and diversification in a New Guinean context. There are numerous studies that investigate language change in New Guinea in the context of a synchronic shift from vernacular to Tok Pisin (e.g. Kulick 1992, Fedden 2012), or changes within Tok Pisin (e.g. Sankoff 1993, Sankoff and Brown 1976, Smeall 1975), but these studies can only obliquely tell us what sociolinguistic pressures may be at work in social contexts outside those of national language use. To the best of my knowledge, there are almost no synchronic studies of language change phenomena within vital vernacular contexts of New Guinea (Schokkin and Lindsey 2018 for an exception).

So what socio-cultural pressures have specifically been proposed for New Guinea, as well as in sociolinguistics more broadly? Laycock (1991) suggests that Melanesian *cultural attitudes* lie at the heart of the multiplicity of languages and dialects in the region. This cultural attitude views language as an *emblem* of distinct group identities (Grace 1981). Laycock's example of how Uisai dialect speakers of Buin (Papuan, South Bougainville) reversed all anaphoric gender agreements to differentiate their community from other Buin speaking communities (Laycock 1991:35) is a well-known one in Papuan linguistics.[9] The observation that language is used for group demarcation purposes has been made throughout ethnographic and linguistic research in New Guinea (see Sankoff 1980c:112-116, Foley 1986:2-9), and is raised as a major candidate of a cultural trait that affects change.

The notion of identity and its role in language variation and change more generally is a mainstay of sociolinguistics. Labov's seminal study of Martha's Vineyard (1972) showed very clearly that identity and cultural orientation was a key factor in explaining why some Vineyarders used *ay* vowels rather than *aw* vowels. The Vineyarders who identified as locals tended to use the central variety of

[9]Evans (pers.comm.) has commented that he has been unable to find examples of this supposed reversal of anaphoric gender agreement in the Buin dictionary (Laycock, 2003). I myself have had a look in the dictionary at the pronominals section and can find no mention of this.

diphthong as a rejection of perceived pressure from the mainland, i.e. they were orienting against mainland pressure. Psycholinguistic research has shown that the desire for social comparison and differentiation are reflected in speech when primed (Bourhis & Giles, 1977). What exactly constitutes and is involved in "expressions of identity", however, is widely discussed. As Mendoza-Denton (2002) points out, variationist sociolinguists have taken three kinds of approaches to the notion of identity. The first approach, which she calls "sociodemographic category-based identity", has been very effective in describing macro-patterns of variation such as gender, class, and socio-economic status. The second approach to identity is a "practice-based identity" which focusses on the "the level of the construction of social relationships" (Mendoza-Denton 2002:486). These include how Le Page and Tabouret-Keller (1985) characterise linguistic behaviour as "acts of identity", which involved group-oriented acts alongside expressions of individual agency and the social construction of identity. The third approach identified by Mendoza-Denton is "practice-based variation", which focusses on variation at the level of the individual (akin to what Eckert (2012) calls "the third wave of variation"). Multiple studies have shown the cross-cultural expression of these various levels of identity (e.g. Mansfield 2015 for Aboriginal Australians (Murrinh-patha, Daly River; Australia), Miller 2013 for the Dane-Z̄aa (Athabaskan, Na-Dene; Canada), Solheim 2009 for Norwegian), and these identities presumably play a role also in New Guinean communities.

In addition to the above mentioned, I believe the kind of identity that is closely linked to the notion of language as emblem is what Hazen (2002) labels *cultural identity*. In his own words, cultural identity is "a sociolinguistic factor that involves how speakers conceive of themselves in relation to their local and larger regional communities." (p.241). The emphasis is on speakers' and communities' *orientation* towards local communities, i.e. how they relate to and sometimes define themselves in opposition to other communities. The notion of cultural orientation thus recognises how a speaker and their community are situated within a larger milieu in addition to the language ecology. Cultural identity as a concept recognises that linguistic behaviours are a "series of acts of identity" as Le Page and Tabouret-Keller (1985:14) put it, where individuals negotiate their personal identities as well as social roles within their social world. These are also the ideas implicit in notions of *audience design* (Bell, 1984), and *style-shifting* (Coupland, 2007) where linguistic behaviours always have one eye towards the wider social world. In this view demographic facts about a community are informative insofar

as we understand speakers' cultural orientations. The argument that language ideology and emblemacity play a major role in language diversification is made for other linguistically diverse regions of the world such as Australia and the Vaupés basin of the Amazon. Regarding Australia, Sutton (1978) concludes for the languages of Western Cape York that "belief-systems regarding language" (p.230) and the "ideology of language as ritual property emblematic of clan identity" (p.231) play a part in dialect maintenance. For the Vaupés, Jackson (1974) suggests that the "possession of a distinct father-language" (p.61) is important because this ultimately identifies who can marry whom; what is known as linguistic exogamy.

An additional shared social characteristic between these linguistically diverse parts of the world are the high levels of multilingualism. Research on multilingualism of small-scale indigenous communities is often discussed in light of language contact with a national or colonial language (Vaughan and Singer 2018:83), which in turn have focussed on the role of multilingual speakers in language convergence and simplification; no doubt because this is the most widely visible phenomenon across the globe today. But there is increasing recognition that language contact can in fact lead to diversification (Mufwene 2001, Trudgill 2011, Braunmüller, Höder, and Kühl 2014, Ellison and Miceli 2017, Evans 2019b). The different outcomes of contact are largely put down to the particularities of the contact situation, and a holistic view of a language ecology (Haugen 1972) would suggest that it is the whole socio-linguistic environment and context that leads to particular language outcomes. Convergence is not an inevitable outcome of contact, since it is the individual multilingual speakers in speech communities, their relationships with other speech communities, and attitudes and beliefs pertaining to languages, that all play a part.

Reports from smaller communities like those in New Guinea often mention the deliberate nature of linguistic control and differentiation by speakers. For example Kulick (1997) mentions how speakers in New Guinean contexts "exaggerate" linguistic differences between themselves and their neighbours (pp.1-2). The extent to which long-term language diversification depends on such deliberate actions, however, is not clear. Thomason (2007) suggests that deliberate manipulation of linguistic features occurs often enough cross-linguistically that they can potentially interfere with the application of the comparative method. The vast majority of these deliberate manipulations (mentioned by Thomason and other fieldworkers) are of lexical and phonological variables; those which are

above the speakers' consciousness. Structural variables below the level of consciousness have been found to be more susceptible to effects such as structural transfer and convergence in contact (Epps 2007, Dench 2001, Mous 1994), which raises the need for nuanced cognitive models that take into account the role of active beliefs on variable production and perception (for a recent attempt, see Campbell-Kibler 2016).

The role of speaker attitudes towards other languages within such multilingual language ecologies has been proposed as a factor in explaining the amount of language diversity in New Guinea. Falling out somewhat from the discussion of population size, low speaker numbers have been suggested as leading to long-term multilingualism across adjoining communities (for the highlands, Salisbury 1962, Sankoff 1980a; for the Sepik, Ferree 2000, c.f. Foley 2000). Recent work on what Vaughan and Singer (2018) call *indigenous multilingualisms* shows similar patterns in small and mobile communities. The Vaupés basin is area with such communities (Chernela 2013, Epps and Michael 2017). A similar observation has been made in relation to Vanuatu, where there are also high levels of language diversity (but only one language family: Oceanic). François (2012) suggests that the principle of "egalitarian multilingualism" encourages and enables the maintenance of cultural differentiation. The observation made by François is that that in parts of Vanuatu there are lots of vernacular languages co-existing, because no particular variety in the area is seen as better, more useful, more prestigious than the others. People speak their own languages, and very often speakers of neighbouring languages have at a minimum comprehension of what is being said, because speakers of these different languages are constantly interacting. This kind of non-production based multilingualism is described by Vaughan and Singer (2018) as "receptive multilingualism". The interest and investigation into the relationship between language change and traditional types of multilingualisms is a recent phenomenon, and is an angle of research that links observations from fieldwork in New Guinea to modern understandings of the role of psycho-physiological and socio-cultural selection pressures in understanding language change.

Another socio-cultural selection pressure suggested as highly relevant in New Guinea is the communication networks within a speech community. The concept of *esoterogeny* mentioned earlier is a type of communication pattern that has been invoked for explaining the difference in linguistic structures between Melanesian and Oceanic languages (Grace 1975, 1980, Ross 2001). According

to Wray and Grace (2007), speakers in esoteric communication interact within a heavily shared-context situation (i.e. in-group oriented communication), where interlocutors are familiar with one another. In contrast, *exoteric* communication involves interaction with strangers where one cannot assume shared knowledge (out-group oriented communication). Wray and Grace's hypothesis is that in the long-run, esoterogeny leads to "complex" languages, while exoterogeny leads to regularity and semantic transparency in a language. The notion of esoterogeny does not say anything directly about rates of language diversification, but helps recognise that communities are often (more-or-less) in- or out-group oriented; the key idea being that social networks impact language change in many ways. This broad idea is shared by Trudgill (2011) where he focusses on "the relative degree of contact vs. isolation" as a factor that produces different rates of linguistic change (p.13).

The importance of social networks for describing language variation and change is a well-established idea in sociolinguistics due to the seminal works by Milroy (1980), and Milroy and Milroy (1985). The term *social network* is used to mean "the informal social relationships contracted by an individual" (Milroy 1980:174). The structure of a network affects how language is used between people and can often account for how variation patterns across a population. The structure of a network is determined by three parameters: the openness, density, and multiplexity of the ties in the network. The openness refers to whether individuals in a speech community have ties to outside the community (open) or not (closed). Density refers to whether individuals in a network know each other. A maximally dense social network consists of everybody knowing each other in the community, while a less-dense network will only have some individuals who know each other in the network. Multiplexity refers to the type of ties individuals have between one another. A tie that is based on more than one level of interaction (e.g. your cousins are also your classmates) is multiplex, while a single level of interaction (e.g. a salesperson-customer relationship in a developed economy) is uniplex. Speech communities with dense, multiplex networks tend to be conservative and norm enforcing (Bowern, 2010) with a strong orientation towards the community vernacular (Milroy and Milroy 1985, Lippi-Green 1989). Because there is a strong tendency for social network structure to be associated with certain linguistic behaviours, there is consensus that changing social network structures can also affect the outcome of language change. Ross (1997) has used what is known about social network structures to infer historical scenarios and

past sociocultural factors in order to explain contemporary linguistic structures in languages of the Pacific.

A perennial discussion in language change studies concerns to what extent rates of change are explainable by purely mechanistic social selection pressure, such as social network structure, frequency of face-to-face interaction, and accommodation to interlocutors (e.g. Trudgill 2004a, 2008). There is some evidence to suggest that supporting a purely mechanistic model is quite challenging, and that cultural selection pressures such as identity do exert more selection pressure. Blythe and Croft (2012a) suggests that setting up a null hypothesis that can be rejected would help answer this question in a principled way. He proposes a neutral selection model based purely on speaker accommodation, and biased selection models where identity factors pressure selection. Modelling work by Baxter, Blythe, Croft, and McKane (2009) tested a neutral selection model against a hypothesis by Trudgill (2004a) that speaker interaction frequency alone was responsible for the formation of New Zealand English (NZE). Baxter et. al. found that their model did not match the data on NZE presented by Trudgill, and reject the neutral selection model in this scenario.

Though not strictly socio-cultural, the "isolation" of New Guinean speech communities have been suggested as relevant to diversification. The rationale was that the challenging terrain of New Guinea inhibited interaction between groups, and that this isolation prevented the convergence of languages over time (c.f. Kulick 1997:1). This notion of isolated communities was dispelled in the face of evidence from anthropology, archaeology, and linguistics, which showed the communities of the New Guinean interior as connected and mobile (e.g. Allen 1982, Macintyre and Young 1982). As Laycock (1982) pointed out, the languages with the largest number of speakers are in fact found in highly isolating areas; there is more diversity in areas of easy terrain and extensive trade (p.33).

Again, while not strictly socio-cultural, the population size of New Guinean speech communities is occasionally suggested a relevant factor in high rates of change. This is presumably because population sizes tend to imply certain kinds of social networks, which in turn imply certain kinds of interaction patterns between individuals. New Guinean speech communities tend to be quite small, and Sankoff (1980a) calculated that out of the 629 Austronesian and non-Austronesian languages she has data on, 52% of these have fewer than 1000 speakers (p.97). But it is also true that there are some languages that have over 30,000

speakers (of which Enga has 110,000 speakers by Sankoff's count). As mentioned earlier when discussing macro-level language evolution, low speaker numbers are thought to affect rates of change, but the correlation between change and population is not well understood at this point in time (Greenhill 2014:574). Based on the comparison of basic vocabulary across Polynesian languages, Bromham, Hua, Fitzpatrick, and Greenhill (2015) found that languages with smaller speaker populations have higher rates of word loss, and lower rates of new word gain, but this does not tell us exactly how this relates to long-term rates of change. As Greenhill (2014) points out, multiple drivers interact in the progression of change; population size is just one variable that is intertwined significantly with other socio-cultural pressures, and comprises only one part of multiple variables involved in language change (p.573).

To summarise the large number of hypotheses pertaining to language change in terms of diversification in New Guinea, there are three broad types of factors which have been suggested. 1) There are the factors pertaining to community configuration, such as demographics, population size, and social network structures. 2) Factors pertaining to socio-cultural pressures, which affect the behaviour of people. The suggested socio-cultural pressures can be largely represented by the terms *cultural identity*, and *language ideologies*. 3) Community configurations and socio-cultural pressures exist in the wider context of a language ecology, where languages and varieties spoken by a variety of groups and individuals form the social landscape within which community configurations are embedded, and sociocultural pressures operate.

1. Community configurations: small populations, coupled with dense multiplex networks, are conducive to rapid language change and differentiation.

2. Socio-cultural pressures: language ideologies of language differentiation, and language as emblems of group identity, encourage people to keep languages distinct.

3. A language ecology with multiple languages of relatively equal social power cultivates language ideologies of egalitarian multilingualism, which creates and maintains multilingual interactions between speakers and groups.

1.4 Methodology: Micro variation leads to Macro Variation

I have explored the various socio-cultural selection pressures which are hypothesised to be conducive to language differentiation, which in theory can scale up to explain global distributions of linguistic diversity. The social selection pressures of social-network configuration with interaction patters (in-group oriented) are seen as key. Cultural selection pressures suggested are those of language and cultural identity, and language ideologies within a multilingual language ecology. In a New Guinean context these are thought, explicitly or implicitly, to fall out of speech community population size. While there are ethnographic descriptions and surveys of New Guinean speech communities that explore these selection pressures, there is no study that takes the speaker and speech community (interactor and population) as the base of a micro-evolutionary process as the source of empirical data. My thesis will present empirical data pertaining to these selection pressures by using multiple methods to interrogate linguistic micro-variations in a New Guinean speech community. In this section I will briefly provide the rationale behind the specific methodologies used in this project.

As the Utterance Selection Model (section 1.2) supposes, it is the mutations (*variations*) in utterances that can potentially lead to a speech community changing their realisation of that particular replicator. This Darwinian idea that small variations can lead to potential language change is found across linguistics. One of the clearest articulations in the twentieth century comes from variationist sociolinguistics, where Weinreich, Labov, and Herzog (1968) state that replicators in competition for selection are two or more grammatical forms representing "two ways of saying the same thing" (p.162). This is the basis from which future change emerges. Likewise Evans (2013) states that "any categorical change is usually preceded by variability either within an individual's speech, or within the norms of the speech community" (p.263). The grammaticalisation framework also implies transitionary stages of variability between two stages, stating that it is a "primarily syntactic, discourse pragmatic phenomenon, to be studied from the point of view of fluid patterns of language use." (Hopper and Traugott 1993: 2)

The approach I will take in this project is a combination of qualitative, quantitative, and descriptive methods of studying sociolinguistic variation.

The qualitative data are ethnographic data with the aim of describing the language ecology, speech community, and language ideologies. Language ecology is the relationship between a language and its social environment (Haugen 1972: 57), so takes into consideration not just the target speech community, but the wider social milieu of the social and linguistic landscape. My qualitative data are based on eight months of fieldwork across the three Nmbo speaking villages of Govav, Bevdvn, and Arovwe, though the majority of my time was based in the village of Bevdvn which is located in between the other two villages. Ethnographic observations are supported by prior works in the area, namely the ethnographies of F.E Williams (1936) and Mary C. Ayres (1983). The impressions of fellow-fieldworkers in the Morehead area are also consulted.

The linguistic descriptive data provides two sources of important data. One, it provides a description of basic aspects of Nmbo grammar so we can comprehend the linguistic phenomena under investigation. Highly structured work such as elicitation of paradigms still exposes variability among speakers, and is a strong reminder of how variability is visible even within highly monitored situations where speakers are conscious of the forms they are producing. The description also partially relies on the ethnographic method of observing and noting occurrences in the speech community. Notes on variable use of linguistic variants were identified during the course of fieldwork, and these have been recorded where relevant in the sketch grammar, as well as in a dedicated section of the appendix (A.19).

The quantitative method comprises a sociophonetics study based on wordlist acoustics, and two corpus-based investigations using natural speech data. The first corpus study investigates word-initial [h]-drop, and the second study investigates an emergent topic marker grammaticalising from a relative clause construction. The sociophonetic study and the [h]-drop study are broadly variationist in method. The topic marker study takes a grammaticalisation approach of studying variability of contexts in which the grammaticalising form is occurring. All three quantitative studies identify a variant replicator, and the linguistic and social environments of their use will be investigated. Linguistic environment refers to factors such as the phonological context, morpho-syntactic behaviour, and discourse context. Social environment refers to social categories associated to speakers who uttered the replicator under investigation, such as the speaker's gender. The specific methodological details of each study are given in the methodology chapter (5) as well as in the chapters of each study.

The three linguistic variables were each chosen for a specific reason. Firstly, the design of this thesis project was to study variation at multiple levels of linguistic structure; hence the choice of one phonetic, one phonetic-phonological, and one morphosyntactic variable. Secondly, the variables needed to be frequent enough that a sufficient number of tokens could feasibly be collected in order to run statistical modelling. Thirdly, the variables should be cross-linguistically common, so that prior works from other speech communities could guide the analysis of the Nmbo results.

	Vowel sociophonetics	**Word initial [h]-drop**	**Topic marker grammaticalisation**
Replicator/ Variant	Monophthongal vowel	Initial [h] words	DEM + COP in the form *gym*
Variability	F1, F2, duration	[h] presence or absence	Decategoralisation, phonetic reduction, discourse usage
Linguistic Environment	Phonetic env.	Phonetic env., lexical effect	Morphological agreement, duration, usage in discourse
Social Environment	Speaker age, gender, village of origin or residence	Speaker age, gender, village of origin or residence	Speaker age, gender, village of origin or residence

Table 1.1: Summary overview of linguistic variables investigated quantitatively in thesis.

1.5 Conclusion and Structure of Thesis

The question of language diversity in New Guinea has been of interest for linguists of all stripes. I have framed this question in terms of a macro-evolutionary model in order to outline how one may tackle the question of why there are so many languages in this part of the world. Macro-evolution raises the parameters of modes and tempo of change as relevant; that is, New Guinean languages are hypothesised to have faster rates of change which leads to faster rates of differentiation. Tempo is affected by mode, and the mode of socio-cultural selection pressures are hypothesised as particularly relevant in New Guinean language change. Many micro-level processes occur and potentially accrue, which may flow on to a macro-scale with changes in linguistic structure over time. Studying micro-variation as part of a larger evolutionary process has been in the do-

main of variationist sociolinguistics, grammaticalisation approaches, and other corpus-based studies of language change. Models of micro-evolution, such as the Utterance Selection Model of language change, place variation, the speaker, and the speech community at the center stage of long-term language change. The modes of socio-cultural features of a speech community are recognised as crucial drivers of replicator selection. Yet there have been few empirical studies of New Guinean speech communities conducted, whose data can be used to extrapolate on historical changes. This thesis aims to fill this gap.

I have identified three main socio-cultural parameters that are proposed as relevant in offering socio-culturally motivated explanations of language diversity in New Guinea. The goal of the thesis is to provide both qualitative and quantitative evidence to support and/or question proposed socio-cultural features of relevance:

1. Community configurations: small populations, coupled with dense multiplex networks, are conducive to rapid language change and differentiation.
2. Socio-cultural pressures: language ideologies of language differentiation, and language as emblems of group identity, encourage people to keep languages distinct.
3. A language ecology with multiple languages of relatively equal social power cultivates language ideologies of egalitarian multilingualism, which creates and maintains multilingual interactions between speakers and groups.

This thesis is organised into three main parts book-ended by the introduction and synthesis. The three parts are Qualitative Descriptions, Linguistic Descriptions, and Quantitative Descriptions.

Part 1 is the qualitative descriptions of Southern New Guinea, and the language ecology of the Morehead area within which the speech community of Nmbo is situated.

The *language ecology chapter* (chapter 2) provides a qualitative ethnographic description of Southern New Guinea and the Morehead area specifically. This chapter provides details on social organisation and interaction patters, which are the details of local socio-cultural pressures that play a role in language variation and change. Data are based on eight months worth of participant observation, and are triangulated by other ethnographic descriptions by anthropologists, linguists, and religious workers who have experience of the area. The picture to

emerge is that of a society of intimates with a small population, and dense and multiplex social networks, but with additional characterisations including a tendency towards in-group orientation and individualism. The speech community is described as embedded in an egalitarian multilingual language ecology where language is significant in culturally orienting against other tribes of the area.

The *Nmbo speech community configuration* chapter (3) provides the specific details of the speech community in addition to the language ecology description. This chapter delineates the speech community, and provides basic demographic facts about the Nmbo speaking villages. The locally relevant social categories of age, gender, and what I call the Other Village (section 3.3.3) will be explained, as these social variables will be used to investigate sociolinguistic patterns of the language data.

Part 2 is the linguistic description section with the *sketch grammar* of Nmbo (chapter 4). The purposes of the sketch grammar are twofold. Firstly, as mentioned earlier, the sketch grammar is an interpretive aid to understanding the linguistic examples used in this thesis. It is a utilitarian grammar without broader discussions on theories on grammar. Secondly, Nmbo is an under-described language, so this grammar can be understood as a documentation sketch reflecting what I currently know about the structure of the language.

Part 3 presents the quantitative studies of variation. Each study investigates a particular linguistic variant in order to identify socio-cultural variables responsible for the patterns of variation

In the *methodology* chapter (chapter 5) I describe how the data were collected and analysed. Two sources of data were created for the quantitative investigations: the Nmbo Wordlist Corpus (NWC), and the Nmbo Sociolinguistic Corpus (NSC) for the naturalistic speech data.

The first quantitative study is a *sociophonetic investigation of Nmbo vowels* (chapter 6). The parameters of vowel height (F1), fronting (F2), and duration are measured. Descriptive and inferential statistics are used to identify macro-social categories that may be responsible for explaining the patterns of vowel realisation.

The second study is a Labovian variationist study of *word initial /h/-drop* (chapter 7) in common nouns. This is the first study in this thesis based on natural speech data. This lexical-phonological variable appears to be below the levels

of consciousness of the Nmbo speech community, despite [h] being a phoneme which non-Nmbo speakers sometimes identify as unique to Nmbo.

The third study is another natural speech corpus-based study, but takes a grammaticalisation approach. The linguistic variable is the *emergence of a topic marker*, formed by an adverbial demonstrative *ge* phonetically merging with a decategorialised copula verb *ym* (chapter 8). This emergent topic marker is based on the DEM+COP constituent losing its morphosyntactic characteristics, resulting in a fossilised form *gym*. This is the sole morphosyntactic variable covered in this thesis, and provides a reference point to understanding the role of different levels of linguistic structure in variation in Nmbo.

The final chapter is the *conclusion* of the thesis (chapter 9). It is a summary and discussion of the qualitative and quantitative findings, which are then contextualised in light of the larger research question concerning language diversity.

Part I

Qualitative Descriptions

Chapter 2

Language Ecology of Southern New Guinea and the Morehead Area

Gänzänam dena ädi kngorotao, är gs ym bä. Yänzu ziva tkmongai, bä yänzu ziva, bä yänzu ziva. - "That's how the sacred man left the people. He would leave him with his own language, and him with his own language, and him with his own language." (Kata Suga, Perceptual Interview 2015)

A common image of Papua New Guinea is a landscape of high mountains and dense jungles, with a "big man culture" (Sahlins, 1963) of wealth accumulation centred around the raising and trading of pigs. This is, of course, a caricature of New Guinean Highlands' societies. New Guinean societies share commonalities with each other in many ways, but Southern New Guinea is noticeably different to the well-known Highlands and islands of Papua New Guinea in terms of its geography, lack of stratified social hierarchies, and the practice of sister exchange as a system of marriage. There is also no evidence of contact with Austronesian languages in the area. These characteristics all have consequences for how we conceptualise and understand socio-cultural pressures as affecting individuals and groups in the area.

The broad goal of this chapter is to present the language ecology of Southern New

Guinea and the Morehead area specifically. Within this over-arching goal, there are two sub-goals. One is to present the wider milieu of SNG as a geographical and cultural area of high linguistic diversity and multilingualism. Since individuals and speech communities operate within their realms of knowledge and experiences of a wider social world, it is crucial to know these socio-cultural details in order to understand and interpret the linguistic results we find later in the thesis. This description of the wider milieu also serves as a demonstration and reminder that "the Nmbo speech community" is a socially and linguistically complex community comprised of multilingual speakers, which departs from reified characterisations of "a language" and its speakers (what Silverstein (2015) calls the distinction between the "language group" and the "speech community"). The notion of *cultural identity* and *orientation* mentioned in the introduction (Hazen 2002) is of particular relevance, i.e. how individuals and groups define themselves in relation to, and opposition to, other communities. Cultural identity refers to the multiple identities held by an individual based on orientation towards other groups. This notion of identity emphasises the negotiations of identity across various social groups of the socio-cultural landscape, according to circumstance and context.

The second goal is to describe the features of Southern New Guinea and Morehead area societies as *societies of intimates*. As mentioned in the introduction, these features are: low levels of adult language contact, high social stability, small size, dense social networks, and large amounts of community-shared information (Trudgill 2011:147). The societies of the Morehead area indeed have these features, but have additional characteristics pertaining to in-group orientation and individualism which are also relevant in understanding peoples behaviours and social interactions.

The Nmbo speech community is more specifically located within an area I refer to as the *Morehead area*. The Morehead *district* is an official administrative label which covers 20 villages that include areas traditionally associated with people who speak Yam languages of the Nambu branch, and also includes the village of Dimsisi which is associated with speakers of Idi, a Pahoturi River language. Since this official designation excludes areas that are still of relevance to the social and linguistic ecology of Nmbo speakers such as the Komnzo (Yam, Tonda Branch) speakers, and Ende (Pahoturi River) speakers, I use the slightly looser of term Morehead *area*.

The Nambu branch area is a dialect chain; a common phenomenon in New Guinea that makes language classification challenging (McElhannon 1971, or what Tryon (1976) calls "language chains" in Vanuatu). As Mühlhäusler (2006) points out, demarcating boundaries then becomes an exercise in establishing arbitrary reference points along the chain (pp.30-31). Furthermore, the prevalence of multilingualism across communities makes moot the classic criterion of "mutual intelligibility" (Laycock 1979). For the purposes of this thesis, however, for ease of exposition I call the varieties within each language family *languages*, with the following acknowledgements. 1) High levels of multilingualism are prevalent in the area, thus mutual intelligibility is not a reliable means to distinguish varieties as languages. 2) Some some varieties are structurally much closer to one another than others. The differences between Nmbo and Nen (a sister language of Nmbo) are, to use the simile by Evans (2012), "like Spanish to Portuguese or German to Dutch." (p.125), while Kerake Nmbo and Yarne Namna/Nambo are much more like dialects (we will discuss these two varieties in-depth later).

I will begin this chapter by describing the physical geography and ecology to set the scene, but also to elucidate how these physical geographic factors affect the culture and social interactions of people in the area (section 2.1). The effects of environment and climate will be illustrated in terms of how subsistence farming creates spaces of dense, kin-based interaction (section 2.2). General community configurations such as population will be outlined, as well as the creation of social connections by marriage and adherence to sections/clan as a social category (section 2.3). The picture of a society of intimates will be made more nuanced by describing people's reservation towards out-groups, combined with a tendency to respect individual autonomy (section 2.4).

Turning then to linguistic matters, I will present what is known concerning the language diversity of the area in terms of language and language families (section 2.5). We will explore how macro social groups are conceived of in terms of an emically recognised relationship between languages, groups of people (tribe), and ultimately place (section 2.5). Finally details about the indigenous multilingual landscape will be presented, which ultimately presents the ideology of egalitarian multilingualism as manifested in the Morehead area (section 2.7).

2.1 Geography and Its Effects on Social Life

Southern New Guinea geographically refers to the lowland area enclosed by the Digul River (Papua Province, Indonesia) to the north-west, the Fly River (Western Province, Papua New Guinea) to the northeast, and the Arafura Sea, Torres Strait, and Coral Sea to the south (Evans, Arka, et al. 2018:641). This area has been variously described as "Trans-Fly" (Williams, 1936), "Southwestern Papua" (Ayres 1983, Hitchcock 2010), "coastal lowlands of Papua New Guinea" Chappell (2005), but these terms limit themselves to the nation state of Papua New Guinea. I will use the term Southern New Guinea (SNG) in recognition that the Indonesian side is geographically contiguous, and will paint a picture of how historically peoples have had loose contact with one another in various way.

The physical geography of SNG displays a striking contrast to the highlands and islands of Papua New Guinea. With elevations of less than 30 metres above sea level, and the distribution of certain flora and fauna, SNG has been observed as strongly resembling tropical northern Australia and adjacent areas (Evans, Arka, et al. 2018:648, Williams 1936:5). As mentioned earlier, it is a relatively new area of land which was formed about 10,000 years ago when the land connection linking New Guinea and Australia was broken by the rising sea levels between the late Pleistocene and early Holocene (Hitchcock 2010:73). The current land was gradually built up by the accumulation of sediment (Evans 2012:112).

There are numerous rivers of varying sizes that drain out to the southern seas. Smaller waterways play a major role in daily life for communities settled away from the coast. Locally important fresh and salt water creeks provide fishing areas. Fishing is done by women in groups, or by solitary men. Children play a variety of games such as wñe kembone[1] and ynda-ynda[2] in creeks and deeper pools.

The ecology and climate have consequences for settlement-subsistence patterns in the area. Ayres (1980:5-6), based on Paijmans, Blake, and Bleeker (1971) gives a detailed and non-technical explanation of the different ecological zones around Rouku village in the Morehead District, identifying four distinct zones. There

[1] A diving game. A broken branch is hidden somewhere in a pool of water. The aim is to find the hidden branch. The person who hid the branch watches on. Nmbo Dictionary (2018) entry wñe kembone.

[2] A game. A person climbs a tree that is growing on the bank of a river. The person will yell out "sawa taka-taka ynda-ynda free!" and walls into the water with legs straight. Nmbo Dictionary (2018) entry ynda-ynda.

is the monsoon rainforest (Nmbo: *orng*) with higher ground and relatively rich lateritic soils.[3] Settlements are usually built in such zones, presumably because of the richness of soils for plant cultivation. There are the clear places (*fan*) with little to no timber, and poorly drained acidic and clayed soils. The *fan* is a hunting space for wallabies, bandicoots, and a variety of birds. Further afield there are the open places (*ka*) which are considered useless land, even for hunting. Finally there are the seasonal swamp lands (*fuvuye*) which are typically timbered with paperbark trees, and swamp floor covered with litter of dead leaves.

These environmental niches transform quite drastically depending on whether the area is in the dry or wet season, and social life shifts accordingly. The climate of the area has been described as "dry tropics" (Ayres 1983 citing Barrau 1968:114) with strongly demarcated seasons. The dry season (*evh yu*) starts around September, with the peak somewhere between October and November. The wet season arrives in late December to early January. Between May and September people are at their freest in terms of physical mobility. Active agricultural food production is at its lowest. The grounds are still too damp to begin burning garden spaces, so people engage in various other activities. Women go fishing in groups as the water ways are still running free. Many church gatherings take place during this time. As the weather warms and months progress towards the onset of the dry season, some families start planning out fencing for their new gardens.

Creeks and rivers dry up during the peak of the dry season in October-November. People get water by one of two ways. One way is to use dug out water holes of about two meters deep. The one in Bevdvn village was made in the 1970s and 80s with the help of simple modern machinery brought in by church groups. The second way is to relocate to areas with fresh water. Villagers in Govav, who are further away from large rivers, walk to larger fresh water rivers further afield during the dry season. At the height of the dry season, villagers often abandon their villages in favour of camping near fresh water or at their garden hamlets. Because people temporarily leave their villages during this time, social interaction is limited mostly with kin or others who share the garden hamlets that are closer to water (see Ayres 1983:155-156 for accounts of Rouku village).

The wet season (*yowqr*, or *yhqr*) starts in December with the arrival of the westerly wind (*bolmbol*). Physical mobility continues to be low during this time of

[3]Soil and rock type containing iron and aluminium

Figure 2.1: Images of environments of the Morehead area. Top row, the *fan* 'savannah', bottom left *orng* 'jungle' with small creek, bottom right shows water catching in June, well-past the wet season.

the year. Rainfall is heavy between January and April. Rivers and creeks flood, and swamp lands become inundated. Döhler measured peak inundation near Rouku village at one metre in 2015 (2016:8). As Williams noted back in the 1930s, however, floods never occasioned any real scarcity (1936:5). Gardens are made on comparatively high grounds so avoid complete ruin.[4] Many animals are hunted and caught in their attempt to reach higher drier grounds. Mosquito numbers explode. One Arovwe villager aged over 50 told me how, in his childhood, people spent most of their time inside smoky houses because mosquitoes are averse to smoke. The introduction of mosquito nets in the area have changed this practice.

[4]This is the case also because of the maturity of the crops in December. They are often developed enough that they can withstand some dampness. The excessive rains of 2017 during March through to August, however, saw many crops rot in the ground as they were not large enough and succumbed to rot.

2.2 Subsistence Farming As Example of Locus of Social Interaction

Societies of the Morehead area are swidden agriculturalists who are dependent on the land. For the remainder of this chapter we will see that much of food production can be, and is, organised around the nuclear family and small units of kin. These garden hamlet groups are the most likely loci of truly unguarded vernacular speech.

The staple crops common in SNG vary across landscapes, ranging from sago in swampy areas to cassava and taro. The staples of the Kerake specifically are cassava for daily life, supplemented with taro and small quantities of sweet potato. Consumption of sago occurs occasionally. The prestige crop, however, is the yam and its many varieties. Nmbo has a rich vocabulary of yam types, and the social standing of an individual (usually a man) revolves around the ability to produce large specimens, and quantities. As William reported, an individual can shame another individual or make them back down in an argument by claiming their opponent's yam harvest is pitiful and therefore does not allow them to make a statement against them (1936:214). Some of my consultants confirmed this stance as applicable today. Yams are, however, labour intensive crops with relatively low calorie yields compared to other tubers; 440 kilojoules per 100 grams compared to cassava with 600kj per 100g by one calculation (Lebot 2009:xvi). Yams have an advantage over cassava and taro, however, since they can be stored up to six months (Asiedu and Sartie 2010:307).[5] The labour intensiveness of yams is notable compared to the other tubers, but the cultivation of this crop can be done by a single individual.

Subsistence agriculture is conducted on a nuclear family basis. Garden plots are owned by a nuclear family, and these family plots are subdivided into rows for individuals: father, mother, and even children receive their own small plot (Williams 1936:215). A single garden plot is cleared to grow the crops necessary to support a single household. Most of the labour is done by members of the nuclear family. Garden plots are made by clearing trees by fire, and some families today use metal axes. The resultant ash from the burning provides fertilisation for the soil. Gardens require clearing every year, and this labour can, in theory, be performed by a single household unit. Some households in Bevdevn indeed did

[5] Although my impression in the Morehead area was that yams can be stored up to about twelve months

operate their gardens as a single household unit with no other households sharing the adjoining areas. It is more common for multiple households to form garden hamlets and to help each other in clearing the garden and other associated tasks. The organisation of these multi-household units seems to reflect the individual preferences of each household. The father of the aforementioned household which gardens alone was described as "a funny fellow" by some of my consultants. This fellow is known in the village as an individual who preferred his personal space, and I did not see him return to his house in the village during the month of September in my second year of fieldwork. There did not seem to be any social penalty for this preference, and we will discuss the respect of individual autonomy as a socio-cultural feature of the Morehead area in a later section (2.4).

Figure 2.2: Images of garden places.

The maintenance of the garden is conducted almost entirely by members of the household, but inter-household co-operation also occurs. The people who help out with garden tasks are either other families of the shared garden hamlet, or sons and sons-in-law of the head of a household. A man will sometimes go and help his wife's family to clear the garden, or to build fencing. This usually

involves travel away form the man's own village to his wife's origin village, since marriage is typically village exogamous. A woman will help at her husband's parents' garden plot, more so if they are elderly.

Garden plots that belong to individual households may then be clustered around garden hamlets. Hamlets are spaces located close to a garden plot, with temporary shelter and food storages. The hamlets are typically occupied by men who are related, since the claim to a tract of land is based on patrilineal descent. I have personally visited two hamlets; one occupied by three brothers, and another larger hamlet that consisted of six families. The six family hamlet is an interesting case, where five of the families were headed by men who all descended from two brothers. The other family consists of a wife descends from one of those brothers, i.e. the male head of the household is kin only though marriage). Hamlets are occupied for as long as the garden plots are productive. When a family moves to a new garden plot, the garden hamlet is left behind. Because garden plots are regularly rotated across a few years, the network of interaction formed with garden hamlet members is also potentially transient. Given, however, that garden hamlet members are often kin this may not be the case.

Subsistence related activity in garden plots and hamlets are, then, are important spaces of social interaction. Social interaction in these places occurs on the bases of kin, or personal relationships such as sharing a garden hamlet. The garden hamlet is described as the most relaxed space of socialising in the Morehead area. Families will often leave villages and camp out at the hamlet during the dry season, and will not return to the village unless there is some social event (such as a church gathering, funeral, or a linguist visiting for fieldwork). Evans and Miller who worked with Bimadbn villagers observed that people are more relaxed with one another when at garden hamlets (pers. comm.). They report more overt joking between people, and conversations are freer compared to when in the village. The garden hamlet is the place that most likely invokes truly unmonitored vernacular speech, and this will be addressed again in the section detailing the demarcation of the speech community (section 3). The lack of guardedness in interaction is undoubtedly due to being with family, but also possibly because immediate kin are the only group of people one is completely safe with. In the following section we will discuss some socio-cultural characteristics of the Morehead area, including the possible role of history and sorcery in fostering esoteric social interactions.

This section has described the geographical lay of the land of SNG and the Morehead area, and how social life around swidden agriculture operates within this environmental niche. We will now move on to other social and cultural characteristics of the Morehead area which are relevant to understanding the social and linguistic landscape of the Nmbo speech community.

2.3 Population, Social Exchange, and Social Organisation

The communities of SNG and the Morehead area have historically been small and dispersed. Williams (1936) estimated about 2000 people in the Morehead area for the year 1926 (p.47). Even the Coastal Marind-Anim, who boast a large-for-the-region estimated population of 7160 in the year 1915 (van. Baal 1966:34), are reported as living in small units; villages and tribes of 500 people (p.45, 51). The 2011 census lists the population of the Morehead District as 11,504 people (National Statistical Office 2013), which is a significant increase from the 1926 figures, but still presents an area with low population density. Hitchcock (2010) calculates the population density of the Morehead area to be approximately 0.5 persons/km^2 (p.75).

As far as we can tell, there has been no formalised trade system of material goods across the area. Williams does not mention anything regarding systematic goods-based trade, and there is little evidence of locally based material trade that has been intentionally maintained over time. This contrasts with the well-known trade routes of the Enga (Meggitt 1974, Ballard 1994). This is not to say material exchanges were completely absent, since Williams noted that variants of the word *sukuva* are used to name tobacco from the Kiwai islands east of the Fly Delta through to Bensbach in the west (1936:424) (in modern Nmbo, it is *skova*).[6] The trading situation in SNG is perhaps much like what Lawrence described for the Madang area along the northern coast of New Guinea, where "individuals conducted their affairs from a purely egocentric standpoint: in terms of constellations of person to person relationships" (1964:28). The traditional importance of establishing and maintaining social relationships through symbolic exchanges over that of material exchanges has been noted in Melanesia (cf. Sil-

[6]Hays (2005) investigates the diffusion pathways of plants in New Guinea, and notes the prevalence of words meaning 'tobacco' formed by the consonant set s-k-b, s-k-b, s-k, and s-b (p.636)

litoe 1998), and so it seems that the trade of material goods is not formalised in SNG. Today there is meagre trade where some men travel from the Morehead area to Sota in West Papua to sell crafts and bird plumes, but this appears to be sporadic.

Social exchange, however, is highly significant in the form of marriage. The Morehead and Bensbach areas have a virilocal and exogamous system of marriage, which take the particular form known as *direct sister exchange*. In its idealised form, two men from opposing sections will exchange their true sisters on marriage. Different communities across SNG have different ways of organising groups and prescribing marriage, and in the Nambu branch language area there is a system with three sections; the Bangu, Sangara, and Mayawa. These section names are identical to what Williams (1936) reported, and span different peoples and language groups from Bimadbn (Nen speaker village) in the east to Rouku (Komnzo village) in the west.

Here is a good moment to clarify the terminology of group organisation used in this thesis: *tribe*, *section*, and *clan*. In the Morehead area, *tribe* is a way of demarcating territorial bounds (Williams 1936:33). Tribal labels such as "Kerake" are used to identify a person's geographical rights, and tribal membership is spoken of as being inherited by patrilineal descent. It is not, however, a system used to specifically trace descent. Cross-cutting this territorial group is the system of *sections*, which by Williams's own admission, is an imperfect term (1936:34).[7] The section system, much like the type *moiety*[8] in broader anthropological literature, prescribes that a person should marry a person from a group outside their own section. The rules are such in the Nmbo area that Sangara and Bangu can marry each other, with complications for the Mayawa since they are further divided into Bangu-Mayawa and Sangara-Mayawa.[9] There appears to be a merger of this three-section system at the Nen and Idi border, where Idi speakers use a combination of the three section system in addition to another clan system (Schokkin pers. comm.). Ende speakers further east only use the clan system (Lindsey pers. comm.). Nama speakers are also reported as not

[7]As Ayres points out, the term "section" conjures up terminology used in the description of Australian four-class systems, but the usage in a SNG context has no bearing on these (p.183).

[8]Derived from the French word *moitié* meaning 'half', moieties are two groups who select marriage partners from each other.

[9]The Mayawa have the most restrictions in the Nmbo area, especially in Govav where Sangara-Mayawa are disallowed from marrying a Sangara from Bevdvn specifically (Williams 1936:123).

making further distinctions beyond the three sections (Siegel pers. comm.). If Nmbo speakers use some kind of clan system, I am not aware of how it works. *Clan*, as used by Williams refers to a totemic group of minor significance; so much so that Williams says that "this system seems to be some extent forgotten [by the people of the Morehead area]." (1936:34). Ayres (1983:189) and Williams (1936:96) both report the existence of totemic groups called *tuarar* which seem to be organised along patrilines. As Ayres notes, this knowledge is secret knowledge and people are unwilling to come forward with this sensitive information (1983:195-6),[10] and since it does not seem particularly pertinent to local concepts of social categorisation I will exclude it from further discussion.[11]

One idealised form of direct sister exchange is the prolonged reciprocal exchange of sisters across generations. This is where a daughter of a sister exchange marriage will marry into her mother's origin village. For example, a Nmbo speaking Kerake woman whose mother hails from a Nen village may marry into that Nen village. Her daughter will then marry a man from her origin village in Kerake lands, and then her granddaughter will marry into the Nen village, and so on. The result is that certain families will have built up deep ties across certain tribal lines over generations. Williams hypothesises that exogamy aids in people entering friendly relations with outside groups "by enabling the exogamous group to bind others to itself" (1936:169). A relationship predicated on direct sister exchange would make the relationship particularly strong. This is reflected by special kinship terms reserved for these relationships in the various Yam languages. In the Nmbo language, two men who have exchanged sisters will call each other *tambre* (*tambera* in Williams 1936:111). The offspring of exchange unions also use a special term *mwite* to refer to one another (figure 2.3).

There is also a strong taboo against uttering the name of affines (i.e. in-law name avoidance). This practice of name avoidance is found widely across societies both within New Guinea (e.g. Stasch 2011 for the Korowai, cf. Simons 1982) and the globe (e.g. Merlan 1997 for Australia , Mitchell 2015, and Treis 2005 for Africa). In the Kerake area both men and women will avoid the names of their in-laws, and will use specific kin terms as address terms, and alternative names as reference terms. Common reference term strategies are biographically significant names

[10]Evans and Döhler report that the *tuarar* is open knowledge in Bimadbn and Rouku (pers. comm.)

[11]This terminological conundrum surrounding what is meant by "clan" raises interesting questions about the suggestion by Stanford (2009a) that clan may be a relevant sociolinguistic variable that patterns variation.

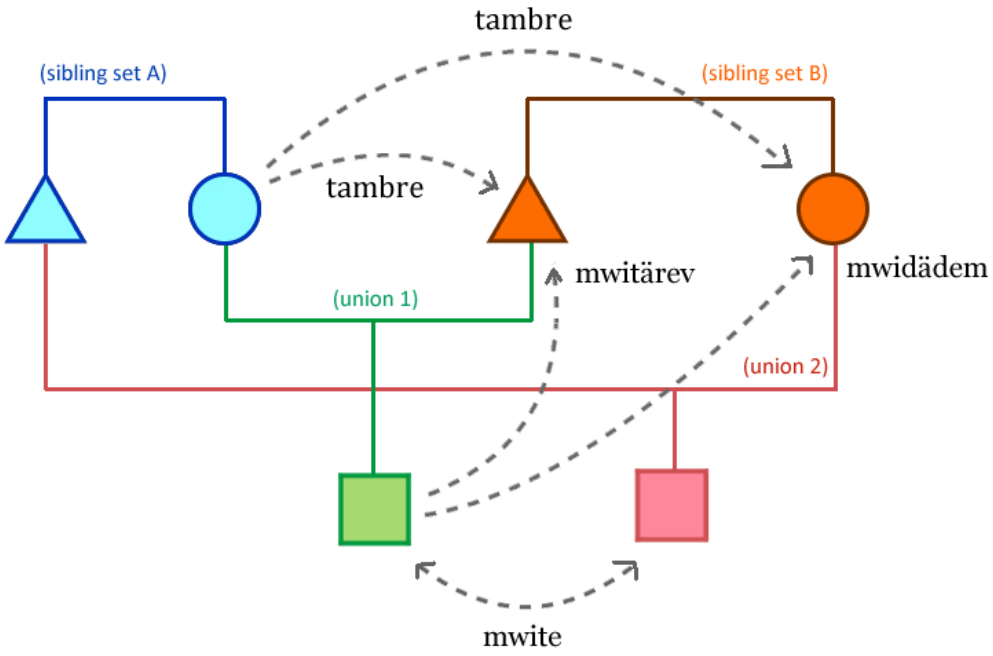

Figure 2.3: Schematic representation of an exchange marriage. The term *tambre* is represented above from the male perspective, but the term is also used by women to refer to their exchange partners.

such as *Kiwäl* 'wind', and indirect referencing such as *Braiyanende är* 'Brian's dad'. As we will see later in the discussion of the Yarne variety (section 3.4), in-law name avoidance was strong enough that speakers chose another form for their object interrogative word 'what'; all in order to avoid uttering a word that sounds like the name of their in-law. This suggests that regardless of whether the union is a true sister exchange or not, the relationship between in-laws is highly valued and sensitive.

The interaction pattern of these low population communities of up to 500 villagers are precisely what Trudgill (2011) describes as *communities of intimates*. These communities consist of dense multiplex social networks of frequent face-to-face interaction. As we will see, however, interaction across each community, whether by village or garden hamlet, is less frequent. The network structure of communities in SNG are best characterised as loosely connected clusters of dense multiplex networks. This characterisation, however, obscures local cultural details which make the relationship between people and groups more nuanced. In the following section we will add the characteristics of out-group avoidance and individualism to this basic description of network structure.

2.4 Out-Group Avoidance and Individualism

The communities of SNG on the one hand demonstrate strong connectivity with one another in terms of marriage exchanges. On the other hand, however, they show varying degrees of out-group avoidance. Secrecy and suspicion of one's neighbours is a well-observed characteristic of New Guinean communities (Wagner 1984, Bercovitch 1989, Stasch 2009, De Vries 2012, Jones 2014), and this characteristic is generally true of Morehead area societies. Williams attributes this to past head-hunting practices, and fear of sorcery (1936:287-91, 355-60), some of which are relevant to understanding Morehead area societies today. Head-hunting has not been a major concern for the peoples of SNG since the early twentieth century, but the impact of this historical practice should be seriously considered. The international border existing between Indonesia and Papua New Guinea was set up in the colonial past due to the "problems" caused by the Marind-Anim head hunters making incursions deep into British colonial territory (Hitchcock 2009).[12] It is unclear whether the Marind-Anim raids of the 1880s through to the 1910s were common-scale or a historical anomaly, and Ayres (1983) rightfully raises questions about how impactful the raids were in terms of population reduction.[13] Given, however, how head-hunting practices are documented for many of the tribes in SNG, we may surmise that it was a common enough occurrence that neighbours were, at a minimum, weary of being too geographically close to one another. Once a cultural convention is set up such that a group which avoids other groups is more likely to survive, a selective process will see the continuation of anti-social practices and beliefs (Boyd 2007:330-331). This results in communities with a strong tendency of in-group focus, i.e. of esoteric communication.

Sorcery suspicions in Morehead to this day may be encouraging the avoidance of out-group members, which correspondingly results in in-group focus. Sorcery in Melanesia has been characterised as having high incidents of accusation levelling towards outsiders, unlike in the African continent where it is common to accuse in-group members (Patterson 1974:139). The responses to witchcraft and sorcery vary across communities in PNG, with physical violence reportedly most common in the highlands of PNG, in some parts of the Sepik, Madang,

[12] As far as the Torres Strait in contemporary Australia, according to Hitchcock (2009)

[13] Ayres does not question the psycho-cultural impact of the raids on the victims of the raids, nor does she question the occurrence of such raids. Ayres raises the question of how much the populations of Morehead in particular were victims of the Marind-Anim specifically (pp.19-21).

and Bougainville (Urame 2015:23). I would not attempt to gauge the level of responses found within the Morehead area regarding sorcery, but Kerake would whisper speculations about events adjacent to their own community, and occasionally within their own. In such cases people will name individuals thought to be responsible for that particular act of sorcery. In my second year of fieldwork, a man whom I worked closely with in the first year delayed greeting me for a few weeks because he felt unwell visiting my host village. His explanation was that a man in my host village had made sorcery against him, and he didn't want to be near this sorcerer. While anecdotal and low key, this small incident demonstrates how belief in maleficia affects interaction between individuals. Sillitoe (1998) describes what he calls the "fearful friends syndrome" in Melanesia, where sorcery fears inhibit secure relationships. "Sorcery encourages demonstrations of amicability for fear of the possible consequences of a serious rift with others, but people are guardedly friendly to those outside a small circle of relatives because underneath their sociable show they are terrified of one another" (p.175). Stasch (2009) describe how the Korowai of West Papua organise many aspects of their daily lives around trying to stay separate from "demons and witches", the latter who live within the Korowai society (p.6). Williams' assessment of sorcery in the Morehead area is that in the main it is "a disruptive force in society" (1936:358) and is in fact highly responsible for social organisation that leads to esoteric modes of interaction:

> "In Keraki society the suspicion does not fall within the limits of the essential social unit, viz. the local section group - and this, in normative terms, may be regarded as a mercy for its members. It is always the outsider who is supposed to make the magical attack, sometimes a member of another village, sometimes of another tribe. But even so modified the consequences of the belief in sorcery are serious enough. The population of the Morehead district is for various reasons scattered, semi-migrant, and lacking in cohesion; and one of these reasons is the belief in sorcery as practised by neighbouring or distant groups. We have seen that it can lead to warfare and the consequent scattering of the people attacked. And at least two of the apparently well-established villages which I found during my earlier visits to the district were later abandoned expressly through fear of sorcerers." (Williams 1936:359)

Another feature of societies in the Morehead area is the high degree to which

individual autonomy is respected. Individuals are held highly responsible for their own actions. The societal characteristic of respecting individual autonomy has been reported for other parts of New Guinea. De Vries (2012) describes the communities of Awyu-Ndumut languages as having a "strong tendency toward political autonomy and egalitarianism" (p.8), which is evidenced by the dispersed nature of living arrangements in the upper part of the Digul River. The following anecdotal story from Govav captures the extent to which people stay clear of what they perceive as other peoples' business. The year of 2015 was particularly dry in Morehead due to the El Niño. While I was visiting the village of Govav, someone lost control of a small fire outside of their house, which resulted in the house burning down. A pineapple garden belonging to a neighbour was caught in the flames and was also lost. While the fire was burning down the accidental arsonist's house, the neighbour did not help to put out the fire. Only when the fires jumped to the neighbour's pineapple garden did he start fighting the fire. All of this was occurring within sight of my consultants and I, as we were working on linguistic transcription. No one offered to help put out the fire. I queried one of my consultants about this, and he responded "it's not my problem". A similar incident happened in the same year in Bevdvn, when a fire burning at one end of the village was fought only by those whose houses were directly threatened by it. While anecdotal, I take both these instances as indicative of a world view where individuals are responsible solely for their immediate social sphere (which might include one's in-laws), but not much further than that. Williams thus states:

> "If you ask a native, 'What will so and so do in the circumstances', he will answer 'Himself!', that is to say, 'He himself will decide. It is not for you and me to anticipate.' The norms of culture, then, are not so strict that deviation is not permitted, and there is often no knowing what a man will do." (Williams 1936: 246)

The overall impression is, then, that section and kin are social categories of high relevance to the cultures of SNG. I have presented socio-cultural characteristics of SNG with an eye towards the Morehead area, which paints a picture of a society of intimates. Morehead societies emphasise social exchange, which is particularly visible in the absence of a formalised economic exchange system. Sections, a social mechanism of grouping people into marriageable groups, is highly salient and important in the Morehead area. The system of direct sister exchange ensures the reciprocal development of relations across groups of people

adjacent to one another in geographic space. Kin relations, especially between in-laws, are particularly important and visible through the practice of in-law name avoidance. Social organisation is that of dense, small communities of low population numbers which are loosely linked to one another through marriage. On top of this, however, there appears to be a latent avoidance of other groups, perhaps inherited from historical fears of head-hunting, and contemporary fears of sorcery. Suspicion of out-groups is coupled with a high level of individual autonomy, which suggest individuals and small groups are likely to maintain strong group identities independent of other groups.

2.5 The Languages of SNG and the Morehead Area

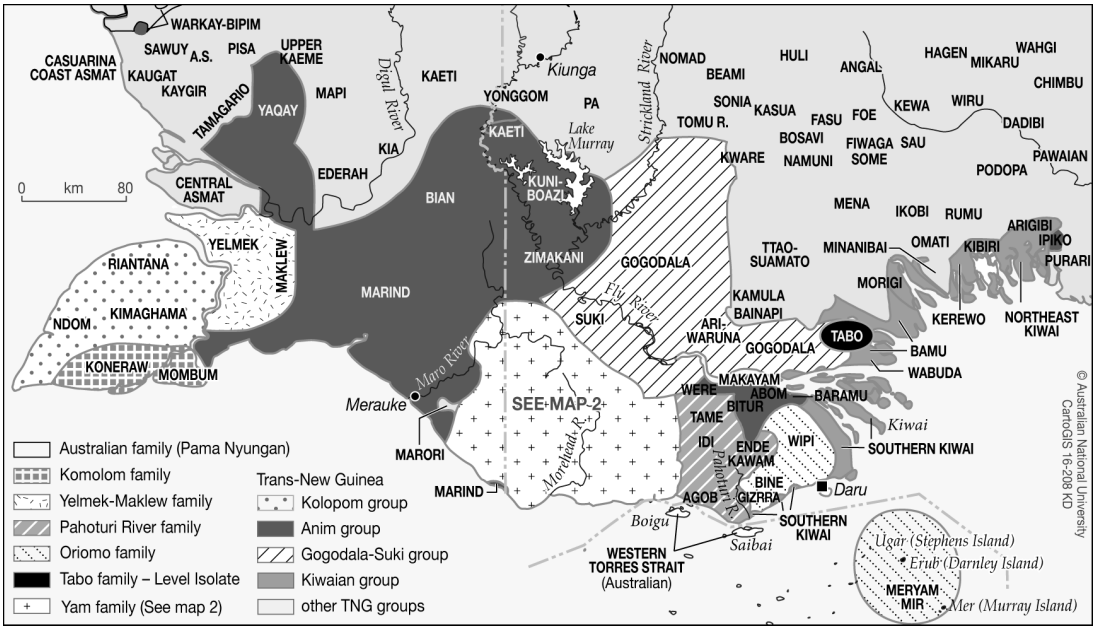

Figure 2.4: Language family distributions across Southern New Guinea. Produced by ANU Carto GIS, from Evans, Arka, et al. (2018):642.

Within the last decade our understanding of the language distributions of SNG have increased tremendously. Papuan languages are still under-described compared to other parts of the world (cf. Hammarström and Nordhoff 2012, Evans and Klamer 2012), but the most recent regional data shows SNG is an area of high phylogenetic and language diversity. The levels of diversity are much higher than earlier estimates such as those made by Wurm (1975) where all the languages west of the Fly River were classified as part of the Trans-New Guinea phylum (p.323). Ross (2005) suggested there were two additional families in

the area, the South-Central Papuan family, and the Eastern Trans-Fly family. A survey conducted by Evans (2012) in the late 2000s showed there were at least five unrelated language families, or maximal clades; Trans-New Guinean, Yelmek-Maklew, Yam (also known as *Morehead-Upper Maro* in Ross 2005:31), Pahoturi River, and Oriomo (Ross's *Eastern Trans-Fly* family). Now there are some forty languages identified, unevenly spread across the five maximal clades. The language diversity of SNG is now seen as second only to those of the Sepik and north coast of New Guinea (Evans, Arka, et al. 2018:641).

Structural disparity is also high. The languages of SNG share typological features across the clades, such as the use of both prefixes and suffixes to index arguments, the large number of inflectional TAM distinctions, a tendency towards multiple exponence, and stem-suppletion of verbs. These are, however, all manifested in various ways by different languages. For example, Nmbo indexes the person and number of the undergoer argument consistently on the leftmost prefix of the verb, while in Coastal Marind the position of the undergoer affix is dependent on the inflectional class of the verb; it can occur either prefixed, suffixed, or infixed to the verb stem (Olsson 2017:27). There are also clear differences in typological features, such as the case system of free pronouns. Yam language have an Ergative-Absolutive systems, while the Pahoturi River languages spoken in the adjacent area to the east have a Nominative-Accusative systems (Evans 2012). Consonant inventories also vary, with Suki (Trans-New Guinea) on the lower side with thirteen, while Idi (Pahoturi River) shows double that number at twenty-six (Evans, Arka, et al. 2018:727, 699).

The Yam family has three branches: Nambu, Tonda, and Yei. Nmbo belongs to the Nambu branch which has at least six identified sister languages. Rueck (2006) identifies six Nambu languages using lexicostatistics among other methods, but more recent reports by Evans, Arka, et al. (2018) suggest that there are more Nambu languages including Nä [næ] and Len spoken in the coastal village of Tais, as well as the nearly extinct Ndre spoken around Arammba (Tonda branch) territory just south of the Suki (Trans-New Guinean, Gogodala-Suki) speaking area. The names used to identify the different Nambu languages reflect lexical differences across the varieties. The names of the languages are based on the word for 'what' in each variety. *Nmbo ym?* in Nmbo translates as "what is it?", *nen ym* in Nen, *nama ym* in Nama and so on. This is akin to having English called *What*, German *Was*, and Japanese *Nani*.

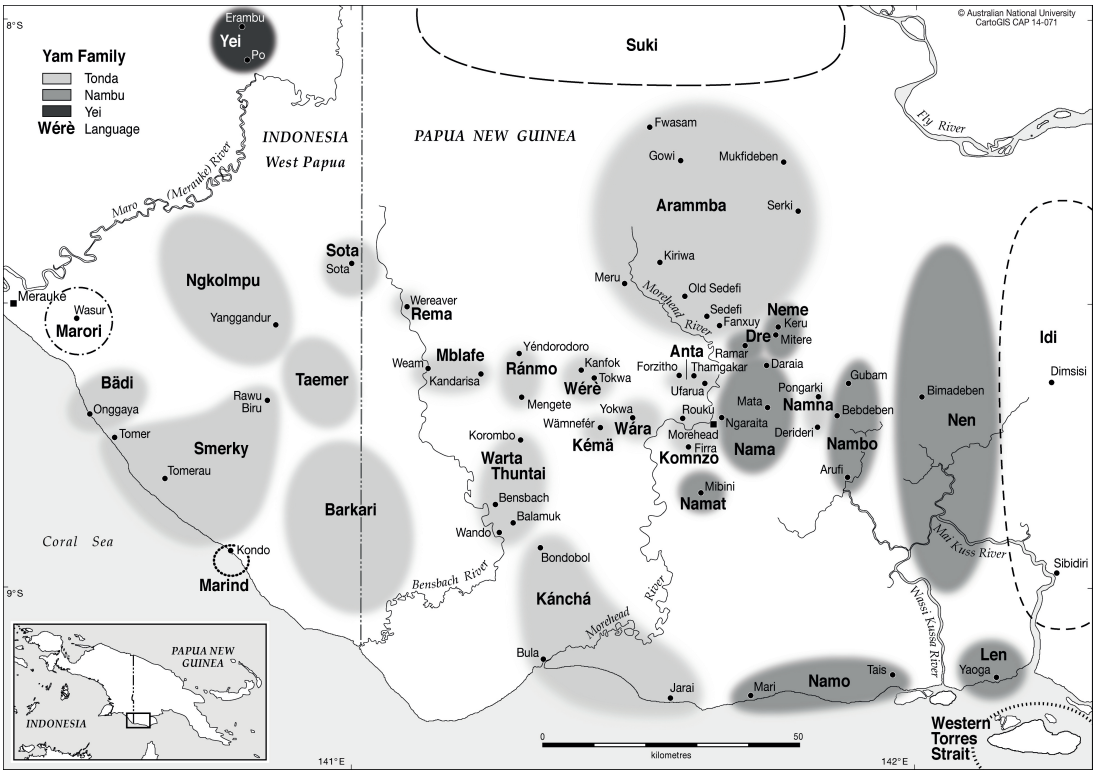

Figure 2.5: Map 2: Details of language distributions around the Papua New Guinea and Indonesian border. Produced by ANU Carto GIS, from Evans, Arka, et al. (2018):645.

The degrees of structural disparity between the Nambu languages is yet to be fully understood. The most well-described differences across the Nambu languages are between Nmbo, Nen (Evans 2012a,2015, 2016, forthcoming) and Nama (Siegel 2014b, 2016, 2017). There are some major differences in the verbal morphology, for example with Nen showing a three-way distinction of undergoer/TAM prefixal forms, while Nmbo and Nama only have a two-way distinction (section 4.5.3). Nen and Nmbo have differing paradigms of pronominal case, with Nmbo making a form distinction between 1ABS (*ynd*), 1sg.ERG (*yndo*), while Nen has collapsed their 1ABS and 1sg.ERG form to *ynd*. Nmbo and Nen show more similarity in lexicon compared to Nmbo and Nama, while Nmbo and Nama share syncretisms in pronominal paradigms as well as case marker form (Ellison, Evans, Kashima, Lindsey, Quinn, Schokkin, Siegel, 2017). More details on the grammatical structures of Nmbo are presented in the sketch grammar (chapter 4). Descriptions of Neme and Nä are still to be conducted.

The two language families of relevance to the Morehead area are the Yam family and the Pahoturi River family. More specifically, we are concerned with the

the Nambu branch languages plus the eastern Tonda branch language Komnzo, and the Pahoturi River language Idi. The eastern-most Nambu language is Nen, whose speakers are found immediately west of Idi. There are quite substantial differences between the Nambu languages and Idi. Nambu languages do not include retroflex consonants as part of their phonological inventory, while Idi does (Evans 2012:136). Nambu languages do not make use of minimally inflecting auxiliary verbs, while Idi does (Schokkin, pers.comm.). Idi also has a productive system of vowel harmony, while Nmbo only shows weak hints of vowel harmony in restricted lexical items that are inflected (section 4.5.2). The Yam and Pahoturi River languages are clearly unrelated, making the Nen and Idi speaking areas a language contact area between two unrelated vernacular languages. The Nmbo speech community is located some ten kilometres west of this high contact area.

2.6 Language and Place and Identity

In the introduction we discussed the importance of identity in understanding language variation and change, and how the notion of cultural identity (Hazen 2002, Le Page and Tabouret-Keller 1985) is particularly relevant in New Guinea. In this section I will argue that the locus of cultural identity in the Morehead area is grounded in the notion of *place* (Ayres, 1983) and *peoples* (tribe). Language, place, and people, are intrinsically linked to one's sense of self and belonging in a group. This concept of belonging as an individual and group will be referred to as *self and group identity.*

The communities of the Morehead area express a cross-culturally common association of self, group, and land. New Guinean societies are described as displaying strong associations between cultural identity and the geographic lands they occupy (Sillitoe 1998, cf. de Lepercavche 1967-1968 for review in context of New Guinea Highlands). Astuti (1995), writing in relation to peoples from western Madagascar, describes this emic concept of self and group identity as a "geodeterminist model of identity". Determinist or not, the emphasis is on culturally meaningful geography and place as forming the foundations of how people view themselves. The association between language, place, and people, is akin to discussions in Merlan (1981) regarding Australian languages and ideologies of "land-language links" (p.139). It has been noted in an Australian context that language and people are associated with culturally meaningful geographic areas

that are often sanctioned in myths and stories (Sutton 1997, Rumsey 1993). This is not to say that communities of the Morehead area and Australia have identical concepts of language, people, and place, or that they may express these concepts in similar ways. The point emphasised here is that self and group identity are inseparable from culturally significant places and languages. This concept of language is in contrast to communities where identities transcend place, such as ethnic identities in an urban context (Fishman 1965, Hoffman and Walker 2010, Cheshire and Fox 2009), class identities (Baranowski 2017, Labov 1966), various life-stages (Book, Jankowski, Konnelly, and Tagliamonte 2018, Bucholtz 1999), life-style and profession (Zhang 2005, Meyerhoff and Walker 2012), and sexual identities (Podesva 2011). Having said so, emic senses of place have also been discussed more broadly by sociolinguists as a way to understand variation in terms other than socially constructed social categories (e.g. Johnstone 2011, Llamas 2007, Eckert 2004).

The different language names are also reflective of different tribal identities. In the Nambu branch there is often a corresponding ethnonym attached to speakers of a language; Nmbo is the language of the Kerake tribe, Nambo/Namna is the language of the Yarne tribe, Nen is the language of the Äkämär tribe, Nama the Tendavi tribe, and so on. The association between language and tribal groups was noted in Williams (1936), though the tribal names used today show some discrepancy with his orthographic representations (e.d. *Wekamara* tribe for the contemporary Äkämär). Ayres (1983) expresses disagreement with Williams's characterisation of language going hand-in-hand with tribe, and her analysis of the Farem people is that place is the singularly most important concept. Ayers described how the Farem people (Komnzo speakers) of Rouku would often describe languages in terms of their geographical location, and that "group of people is not what is referred to in group definition, but a place is." (1983:134) Location and language are strongly associated, but my impressions are in line with Williams, and that tribe is a significant social category of identity. It was common for Nmbo speakers to say "Nen is the language of the Äkämär people", or to talk about the "Gunduma people" (Ayres's Nggunduma people) as speakers of Idi. Sillitoe notes from his experience in the Western Highlands of New Guinea that people identify descent groups geographically with particular locales (1998:142), meaning that the social location of any individual as part of a group that transcends time (i.e. part of a tribe, clan, moiety) is necessarily in part geographically defined.

The association of language, people, and place in the Morehead area is mythically sanctioned. One episode in the cosmogenesis of the Morehead area describes how the land was populated by people who spoke different languages. Each language group tells the story from their point of view, but the key events of the myth are shared in the various tellings across the Morehead area. The episode begins when the primordial *gänzän* (*gainjan* in Williams 1936) hears human voices in the trunk of a *sakr* palm tree. The following excerpt is from Williams:

> "It was to satisfy his curiosity that he took his axe and felled the mysterious palm, so that by bending his ear to the trunk as it lay on the ground he might be better able to distinguish the sounds. Near the upper extremity he found them mere babel, and when he split or lopped the trunk at this point there swarmed out men who spoke the unintelligible language of the Gambadi [Tonda language speakers]..." (Williams 1936:54-55)

Different tellers will choose different languages as this first unintelligible language. A recent re-enactment by Nen speakers in Bimadbn had the upper-most language in the trunk as English. The speaker chooses a language that is geographically and socially distant, but recognisable as a human language. In the Williams version, the *gänzän* then continues chopping off bits of the *sakr* palm, until he hears "namma, namma, namma" which he recognises as somewhat comprehensible, and sets free the Tendavi tribe (Nama speakers) from the trunk. He does this again when he hears "nemati" and frees the Moive tribe (Neme speakers), then "neni" and frees the Wekamara tribe (Äkämär, Nen speakers), and so on:

> "...until, reaching the very butt at last, he heard his own familiar *Numbo*? and so set free the Keraki proper. As they flocked out they threatened Gainjan with their bows and arrows, as each tribe had done before, but their hostility melted when he hailed them as his children and entreated them to speak and gratify his ear with syllables he could easily understand. After this, and before his own ascension into the sky, Gainjan dispatched the peoples and tribes to their present homes, and they very naturally carried their languages and their nicknames with them." (Williams 1936:55)

The crux of the story reveals the teller's language as the one the *gänzän* recognises as his own. Williams was told this story by Nmbo speakers, and so the people

at the base of the tree are revealed as Nmbo speakers. The Nen speaker re-enactment had people streaming out of the felled trunk, with the final person revealed as a Nen speaker. But as the *gänzän* recognised the similarities between the many languages, the myth speaks of a shared link between the languages that are now found across the Morehead area. The *sakr* story gives a folk linguistic explanation of how the Morehead area ended up with so many languages, and also intrinsically situates the languages in geographic space.

The social categories of high importance in the Morehead area are thus fundamentally based on a cosmologically sanctioned association of language, people, and place. Ayres identifies two main social categories of importance in the Morehead area, both of which are based on the importance of place. One of these categories is section, which we have mentioned in brief earlier (section 2.3). The other major social categories of salience are what Ayres calls "dialect group" (1983:133), which we have been calling "languages" and "varieties".

The intrinsic link between language and place is visible in the ways people interchangeably use language names with place names when talking about locations or groups of people. When speaker LS, a middle aged man from Bevdvn, was asked whether he spoke his mother's language Nama as a child, he said no. In his response he identifies the Nama language as being of a place which has no bearing on his origin:

> *Yao. Mnamn, yao ynd kt nglamn* ***Nama zi kndn****.*
>
> "No. Because, I was not born there in the **Nama language area/place**."
>
> (Translated by EK, WSEK1-B20170627-03SmakoSL03MQ, 00:02:37.500 - 00:02:40.163)

Place names can also be used interchangeably with language names. For example Nmbo speakers would often identify languages by the village names when asked about language repertoires. The following are a few typical examples:

> ***Bimadevn zie*** *ädi nngowavtan däv.*
>
> "I speak **with the Bimadbn language** to them then [i.e. when they come to our village]"
>
> (Translated by EK, WSEK1-G20170617-01Alqi02MQ, 00:16:54.929 - 00:16:57.409)

> ***Keru*** *ä näyäretan. Ä nowavtan.*
>
> "**Keru** [the Neme language] I hear it. I speak it."
>
> (Translated by EK, WSEK1-B20170627-03SmakoSL03MQ, 00:08:41.718 - 00:08:43.570)

Language inevitably becomes an emblem of one's identity given the tripartite link between language, people, and place. Languages are a crucial way of situating an individual in the geographic and social landscape of the Morehead area. Ayres emphatically states that "The importance of dialect differences cannot be overstated. They manifest group identity" (1983:330). A person is identified as belonging to a certain group because they speak a particular language, and so language is a primary means of categorising people in the social landscape of the area.

2.7 Indigenous Multilingualism and Egalitarian Multilingualism in the Morehead Area

The large number of languages in SNG and the Morehead area suggest that the communities and speakers are highly multilingual. Indeed, multilingualism is a key characteristic of the Morehead language ecology, and is crucial to understanding the sociolinguistics of the speech communities in the area. The term *egalitarian multilingualism* as used by (François, 2012) is apt for describing the situation in the Morehead area. The Morehead area has non-hierarchical forms of multilingualism between small-scale vernaculars. Lüpke (2016) identifies four types of vernacular multilingualism found in small-scale indigenous communities, two of which are applicable to the Morehead area. Like Lüpke's summary of Northern Australia (summarising work by Singer and Harris 2016), the communities of the Morehead area show a combination of *reciprocal* and *receptive* multilingualism. I will present to the best of my knowledge how these two types of multilingualism are manifested in the Morehead area, and how they form part of an egalitarian multilingual language ecology.

Reciprocal multilingualism describes a societal context where individuals, or an entire community, speak the languages of the neighbouring areas. The beliefs that arises from reciprocal multilingual contexts are the prizing of multilingualism and multilingual speakers, regardless of the exact repertoire in languages. The belief in the value of multilingualism is shared across the community, which encourages reciprocity in production and/or comprehension of different languages in a given area (Lüpke 2016: 51). Situations of reciprocal multilingualism have been observed for the Siane speakers (Trans-New Guinean; Papua New Guinea) of the Eastern Highlands (Salisbury 1962), and is also found in the Morehead area.

One of the beliefs and practices associated with reciprocal multilingualism concerns the use of multiple languages as an expression of socio-political clout, important in the formation of alliances across various groups (cf. Lüpke 2018 for North-western Cameroonian examples). In the Morehead area, the expression of socio-political clout through the display of one's multilingual repertoires is done in venues of public speaking. The display of individual repertoires can occur at profane village-level gatherings such as at feasts,[14] but can also occur in spiritual spaces such as during church services. On my last field trip I attended a church service in the village of Arovwe where a young woman from the Nen-speaking Bimabdn village addressed the mostly Nmbo speaking congregation in both Nen and Nmbo.[15] By orating in her own language as well as in the language of her congregation, the woman was positioning herself as a figure who has the authority to publicly talk about esoteric Christian knowledge. My analysis of this Nen speaker is based on the observation by Ayres (1983) that public speech concerning stories and knowledge is a display of authority in the Morehead Area (pp.335-337). The act of publicly speaking is in itself an expression of socio-political power, so the young Nen-speaking woman's expression of Christian knowledge in both Nen and Nmbo can be viewed as a way of indicating spiritual authority to a community outside her own Nen community.

The ability to use language in socio-political performance is only possible because people in the Morehead area practice reciprocal multilingualism throughout their lives. Nmbo speakers report the ability to speak their mother's language, using the Nmbo expression *bä nowavtan* "I [will] speak it". The ability to speak one's mother's language is due to the exposure and use of the language throughout childhood. Following the practice of place exogamy, a woman will often marry into a place where a different language is spoken. She will, however, continue to speak her language to her children. The following excerpt shows how an out-marrying woman typically speaks about the different languages used in the household. Yayam, who married in to Bimadbn two decades ago, explains how she speaks many different languages to her children. Her repertoire unusually includes Gogodala (Trans-New Guinea; Gogodala) which she learnt during her

[14]Lindsey (2019) describes the *kawa* public oration practice in the Ende speaking village of Limol in SNG, but does not mention multilingualism as a particular feature of the practice (pp.232-3)

[15]My ability to discern Nen is limited to identifying the frequent use of the /p/ phoneme, which in Nmbo occurs only in highly restricted environments. I confirmed my impressions with my consultants after that she was using both Nen and Nmbo in her address, although I did not ask whether other languages were also used.

attendance at a Bible school in Balimo:

> LM: *Bm na zi ä nowavt, bende toge-togeavevav?*
> YS: *Ynd Nen zi nowavtan, Nmbo zi. Amande zi ä nowavtan... ämb evhon... Yna tnamsan, Gogodala zi... kt ymarengrwn. Kanma mwyat wm, bä nngowavtan.*
>
> LM: What language do you speak to your children?
> YS: I speak the Nen language, and the Nmbo language. I will speak mother's language... some times... I used the Gogodala language. I lived there a long time ago [in Balimo]. If I know the language, I'll speak it. [lit. Wherever I know, I will speak a lot.] "
>
> Translation by BA, minor edits by EK. WSEK1-B20170720-02Yayam02MQhm, 00:09:12.874 - 00:09:29.045)

Another aspect that sustains reciprocal multilingualism is the high level of acceptance towards speaking other languages. An individual's ability to speak multiple languages is fostered and actively supported by peoples across multiple communities. Speaker YS reports how in her early days of marriage both her husband and other women of Bimadbn taught her Nen words by translating them into Nmbo:

> *Ge tnmarengrwn ämbru mngon yände Nen zie ge qmndetao, ämb tngawayamngotawn, ämb yao. Ädi qawaverngai. Dena däv ge kä äyäräyär ge kovrotawn nmbo ymndet, qavtangai de, "yao dena."*
>
> "When we stayed together in one house and he happened to talk to me in his own Nen dialect, I could understand some bits, some not. He translated to me. Like, when I had doubts about what he was saying, he would say to me "no, it's like this." "
>
> Translation by BA, minor edits by EK. WSEK1-B20170721-01Yayam09Finalh5, 00:08:40.435 - 00:09:00.000
>
> *Ynan dena tmndongayn, "Krava, tkr ym bende zi. Nmbo ymn ye." Ynan ämb äkämär zi mna tngmaro tekr. Gnoso ävrät. Ädi qawavwenao. A Dämi gs ym, ymo qmdongai yao dena mnan bä Nmbo zi... Ymo qngavngongai "dena nowavt". Är giya yngyäreta dena äyngaowavt yve tekr zi.*
>
> "I use to tell him like this, "Oh gosh! Your language is hard. What

is this." Some Äkämär [Nen] words, you know, they were hard. Now it's easy. He explained to me. And Dämi she would tell me, because she knew Nmbo. She would tell me, "He is saying like this. When you hear people this is how they are speaking their hard words." "

Translation by BA, major edits by EK. WSEK1-B20170721-01Yayam09Finalh5, 00:09:02.572 - 00:09:26.200

The reciprocal multilingual practices in the Morehead area is not gendered, in the sense that both men and women learn the languages of their mothers. This is in contrast to northern New Guinea where multilingualism has been reported as being a male affair, such as by Conrad (1978) for the Iwam (Sepik), and Litteral (1978) for the Anggor (Sengai), both from the upper Sepik river region. Ayres (1983) claims that mothers in the Morehead area speak their varieties only to their daughters (p.180), but to my knowledge this is not the case. There are numerous examples as captured in the following excerpt, where a man reports how he spoke to his mother in her language Neme:

Kerumn är mrz tmaro. Neme zi kowavtao e~, noroi yna Neme zi, nmbo zi wnde kowavtao, ämb mure tawanongayn mnan, ynd mwyat de nmndoyn. Amande zi.

"She [my mother] was a girl from Keru. She would speak in Neme until~, she left the Neme language and was speaking in Nmbo zi. I would answer her just a little bit because I knew it already. Mother's language. "

(Translation BA, minor edits by EK, WSEK1-G20170707-03MY02MQhm, 00:13:59.044 - 00:14:04.806

The repertoires of Morehead area people include other languages in addition to their mother's language. An individual may know anywhere between five and nine languages with varying skill levels of production and comprehension (see appendix for language repertoires of speakers in the Nmbo Sociolinguistic Corpus). Nmbo speakers distinguish differing levels of multilingual ability with a variety of expressions, ranging from *ynd nowavtan* "I speak [the language]", to *näyäretan* "I can hear [the language]", to *dodo fivi* "no good at all". The "I speak it" can be further clarified by distinguish between the use of *avrat zi got* "easy words" such as greetings, and a more thorough knowledge of the lexicon and its uses. These emic distinctions show a sensitivity of metalinguistic awareness, and also hint at what locally constitutes a true speaker of a language; i.e. the

ability to speak a language with all its linguistic and social intricacies. Table 2.1 shows some common expressions used by Nmbo speakers to describe the ranges of linguistic ability. At this stage of fieldwork it is unfortunately difficult to gauge how much modesty or self inflation may be imbued in these self reports. I have recordings of speakers giving demonstrations of their ability to speak another language, and the speakers' willingness to go on the record should be interpreted as significant. Speaking accurately and correctly is highly prized in the Morehead area, regardless of whether you are speaking your own language or another's (see also Ayres 1983:331).

	Nmbo	English Translation
Production: high	Ynd nowavtan.	'I speak it.'
	Mwyat wm.	'I know it.'
		'I know some words.'
	Yao tombae fivi	'Not long words.'
		'Only easy words.'
Production: low		'Can't speak hard words.'
Simple production	Nngangotan.	'I can answer.'
Comprehension	Ynd näyäretan.	'I can hear it.'
Non-comprehension	Mavna wm.	'I don't know.'
	Dodo fivi wm.	'I'm truly ignorant.'

Table 2.1: A non-exhaustive list of common Nmbo expressions to describe degrees of production competence, and comprehension. The ordering of these concepts are not to be interpreted as strictly linear.

While people in the Morehead area are highly capable of speaking other languages, they express a preference for *receptive multilingual* interactions. Receptive multilingualism refers to a context where speakers use their respective languages of choice while speaking to one another (ten Thije and Zeevaert 2007:1, Singer 2018:102). It is a type of multilingualism that is found in many parts of the world across differing societies (e.g. Sutton 1978, Evans 2003, Singer and Harris 2016, Vaughan 2018 for northern Australia, Campbell and Grondona 2010, Epps 2018 for the Vaupés Basin, Verschik 2012 for Finnish, Estonian and Russian, ten Thije and Zeevaert 2007 for business contexts in continental and northern Europe), and has been observed in other communities of the Morehead area such as between Nen and Idi speakers (Evans 2012:119), and in the village of Rouku (Döhler 2018:34-35).

The following excerpts from interviews about multilingualism show how Nmbo speakers describe receptive multilingual interactions. Note the expressions indicating the speaker's linguistic ability of the non-Nmbo language being discussed:

> *Dena gāvut gea wngm Bimadvnat, bN̈en zie gea nowavtat, ynd tanzo zie ä nowavtan, bä bä näyäret, ynd bä näyaäretan yände zi.*
>
> "It's like this, if I go to Bimadbn, if they will speak in Nen, I will speak with my own Nmbo zi, he will hear [understand it], and I will hear his language."
>
> (Translation EK, WSEK1-G20170707-03MY02MQhm, 00:13:59.044 - 00:14:04.806

> *Bava bä yänzu zi kowavtao, yndo kiyäretawn.*
>
> "Bava [mother's brother] spoke his own language, I listened to him."
>
> (Translation EK. WSEK1-G20170728-02Masa02MQhm, 00:06:26.999 - 00:06:29.661)

What does not appear to be a common type of multilingualism in SNG is that of *passive multilingualism*. Lüpke described this more as an ideology rather than that a type of behaviour, which espouse a maximal separation of codes (2016:57). Communities of the Vaupés region of the Amazon have been described in such a way by (Jackson 1974,Gomez-Imbert 1996, Chernela 2013), as well as the Sui in South-Western China (Stanford, 2007). In these situations, women will marry into a community that does not speak her own language, and she will continue to exclusively speak her own variety for the course of her lifetime. She will comprehend her husband's village's language but will not speak it herself. Children acquire their mother's variety early in life, but will grow out of it and speak their patrilect exclusively. This is clearly not the case in the Morehead area, since Nmbo speakers of both genders report speaking their mother's languages. There is passive multilingualism in the sense that some individuals have comprehension without production of a language, but the ideological requirement of linguistic exclusivity is not a strong characteristic of Morehead multilingualism.

The use of languages other than one's own, however, are socially and culturally motivated. The linguistic etiquette of the Morehead area is that one should speak one's own language, but it is also polite to speak in the language of others as a sign of good-will, social bonding, and reinforcement of pre-existing relationships. In the two examples below, both speakers are referring to interaction with kin. I unfortunately do not have a good gauge of how common accommodation is when

speaking to non-kin. In the first excerpt a woman in her forties, Alqi, explains how she would speak her mother's language Neme to her maternal uncles when they visited. In the second excerpt, Gima, a man in his mid-thirties explains how he accommodates to his mother's preferred language Nen, which demonstrates not only a pragmatism concerning ease of communication but also a sensitivity to his mother's needs:

> *Amande zi däv nowavtan dena, amande hakr ynano ge ynm. Ge nnarendat. Ymova däv nowavtan.*
>
> "I spoke mother's language when mother's brothers came. They would come. Then I speak with them."
>
> Translation EK with TT and RM. WSEK1-G20170617-01Alqi02MQ, 00:08:24.803 - 00:08:31.075

> *Amava ziwt gymn... Ama dena na krtayan, endneyan gea kungenga, amande mende wavroh gea tavngon, ok... Amae däv yänzu zie bä yawangotan dena, na gea kowavtanga nuwt... o bs tiaht... amae däv yänzu zie bä... yawangotan. [...] E dävandru dena, amam yänzu mende nmbo gea qatlinga o gea wamdeta... ymova kt Nen zie bä nowavm.*
>
> "I use my mother's language with her, ok... like, if mother has some kind of problem, like she should become sick, I will do what mother wants, ok... I will answer mother in her own language, if she requests for example water....or lighting a fire... then I will use the language [Nen] with her. [...] Yes, when mother needs something, she will report to me what she needs, or she will tell me... that's when the two of us will speak with the Nen language."
>
> Translation GZ and RG with minor edits by EK, G20170622-01GimaGZ-02MQhm, 00:09:18.960 - 00:09:50.240)

Implicit in the reciprocal-receptive multilingualism of the Morehead area is the belief that all languages are socially equal, and no language is viewed as more valuable, or more prestigious. The belief in language equality is best described by the term *egalitarian multilingualism* (François 2012, 2011). The term "egalitarian bilingualism" was first used by Haudricourt (1961) in relation to pre-colonial New Caledonia, but François's emphasis on *multilingualism* is more apt for describing the situation in the Morehead area. In a language ecology with egalitarian multilingual ideals, there is no pressure for speakers or communities to align their

language to that of another one. I have already suggested that communities of northern Australia (Vaughan 2018, Evans 2011, Sutton 1997) and the Vaupés basin (Epps 2018, Epps and Salanova 2013) hold egalitarian multilingual beliefs, and egalitarian multilingual beliefs are also found in American Indian communities of Central California (Kroskrity, 2018). Sankoff's hypotheses on pre-colonial New Guinean multilingualism imagines a similar situation of balanced multilingualism between groups (Sankoff 1980b:13), and this is true of the contemporary Morehead area. François (2012) suggests that the egalitarian attitude towards different languages reflects the egalitarian nature of political power within and across various communities, i.e. a headless society with little social stratification compared to the politically more integrated societies of Polynesia (p.86).

Ideals of egalitarian multilingualism have been argued to be a necessary component in explaining language diversification. Receptive multilingualism which entails egalitarian multilingual ideals, is hypothesised by Singer (2018) as conducive to identity indexation because there is no social pressure to conform one's speech to the language of the other (p.104). François suggests that the twin forces of socially emblematic differentiation, and wide-spread contact between languages, can only be balanced by the willingness to learn neighbouring languages (p.93). We have see this willingness to learn other languages by Nmbo speakers, where children and adults will learn to speak other languages throughout the course of their lives. The willingness to speak other languages is socially motivated by language etiquettes that encourage the speaking of interlocutor languages in appropriate circumstances.

In this section I have described the multilingual ecology from the perspective of Nmbo speakers. The two types of multilingualism are reciprocal and receptive types that entail an ideal of egalitarian multilingualism. The indigenous multilingual ecology of the Morehead area, with its beliefs and practices, make the maintenance of language diversity possible.

2.8 Conclusion

The purpose of this chapter has been to present the wider social world within which the Nmbo speech community is located.

I have set the geographic, socio-cultural, and linguistic stage of SNG broadly, and Morehead more specifically. SNG has a relatively homogeneous geographic profile

with pockets of tight, dense network communities centred around kin. The lack of formalised trade networks suggests that historically there was little opportunity for frequent interaction with peoples from radically different locales or groups. The relationship between individuals is established and enforced less by material goods, but more by social exchange. The strongest form of relationship is based on marriage, and direct sister exchange as found in the Morehead area forms a particularly strong relationship between individuals and groups. All unions beget strong affect and obligations across linages, regardless of whether they fulfil the criterion of a direct sister exchange. At the same time, there is a tendency to regard groups outside kin with suspicion, possibly due to the history of head-hunting, and beliefs in sorcery; the latter of which is prevalent to this day. The interaction style in such groups is ultimately in-group oriented; esoteric in the sense of Thurston (1992), and Wray and Grace (2007), where people are interacting within a heavily shared-context situation with in-group members. All of this results is a vast network across the entire area of SNG which has a high level of language diversity, and multilingualism between vernaculars is a mainstay of daily life.

This chapter has also shown how the high levels of linguistic diversity are reflected in the high degree of multilingualism in the Morehead area. Cultural identity in the Morehead area is marked by people's languages, which are also intrinsically linked to concepts of place. Reciprocal and receptive multilingual practices and beliefs make languages major emblems of group identity, since the requirement of language conformity is eliminated. The willingness to learn multiple languages is socially motivated as ways of signalling socio-political clout, but also as a way of strengthening social bonds between kin, and possibly other groups. The two types of multilingualism also give rise to the ideals of egalitarian multilingualism. Egalitarian multilingualism fosters the ability for multiple languages to co-exist, contributing to the maintenance of language diversity in the Morehead area.

Chapter 3

The Nmbo Speech Community

Ynd ge wm Bevdvn hambamn är ädi wm. Ynd tvenzo yun ädi ynm, tvende bandan ädi yna giym, tanzo bandan ädi wmarengr. - "I who am here, am a man from Bevdvn village. We are staying in our own land here, I'm staying in my own land." (Yavs Zoga, 'My Village')

Nmbo (pronounced /nɐm.bo/, glottocode NAMB1293, ISO-639-3 code ncm) is one of the many languages spoken in the administrative area of the Morehead District of the Southern-Fly Area of Western Province. It is a vital language, and is acquired from birth by children as one of the many languages that form a Kerake child's linguistic environment.

The notion of the speech community is notoriously problematised in sociolinguistics (Schilling 2013:18-28, c.f. Patrick 2002:573-4 for a large list of sociolinguistic studies that use the notion in conflicting ways). Part of the reason appears to be a confounding of the notion of "speech community" as a locally meaningful (or not) emic concept, and "speech community" as an academic and methodological construct for empirical analysis. The collapsing of the two concepts seems to occur in part because the speech community as a methodological construct is partially informed by emic criteria. In this chapter I will delineate the Nmbo speech community for methodological purposes, using a classic definition proposed by

Labov (2007).

I interpret Labov's definitions to fit the cultural and linguistic ecology of the Morehead area (section 3.1), the background details to which I have just given in chapter 2. As Tagliamonte (2012) points out, the dimensions that define a speech community are locally particular; they must be adapted to fit different field sites, and research questions (p.101). I will then present specific demographic facts about the Nmbo speech community (section 3.2), and then provide descriptions on locally relevant social categories of age, gender, and what I call *the Other Village*. In the final section I will describe the situation of the Yarne variety of Nambo/Namna as an act of sociolinguistic documentation, despite the Yarne variety unfortunately not featuring as a central component of this thesis.

3.1 Delineating the Nmbo Speech Community

The speech community is defined by Labov (2007) as having three dimensions: 1) well-defined limits, 2) a common structural base, and 3) a unified set of sociolinguistic norms (p.347). We have already discussed some common sociolinguistic norms concerning language and identity in the Morehead area in the previous chapter, so those will not be repeated here. In this chapter I will delineate the Nmbo speech community focussing on the first two parameters as a guideline.

The Nmbo speech community is well-defined in local ideology as the language spoken by the Kerake. The Nmbo speech community can also be practically limited by identifying people who speak Nmbo to varying degrees of competence. As explored in chapter 2 (section 2.7), a person's ability to speak Nmbo can range between competence in production as a dominant language, through to selectively speaking it as part of one's extended linguistic repertoire. The Kerake who originate from Nmbo places are the *licensed speakers* of Nmbo, who are sanctioned by the local world view as the rightful speakers of the language (section 2.6). The licensed speakers are the people who have a claim on the language, and whose identities are manifested as speakers of the language.

The Nmbo speech community can also be defined in geographically bounded terms. Nmbo speakers are subsistence farmers who live off the land, and spend most of their lives within the geographic range of villages and garden hamlets (section 2.1). There are no significant diaspora of Nmbo speakers outside the

Morehead area,[1] so the the speech community can be delimited relatively unproblematically to the physical confines of the Morehead area. The isogloss representing the geographic range of Nmbo is a good approximation of the geographic spread of the majority of Nmbo speakers.

Village is the most concrete way the concept of place is manifested in the the Nmbo speech community. Villages are a relatively new mode of settlement in the Morehead area (Ayres 1983:10-11), but they are important spaces of social life and identity in contemporary life (p.37). Villages are the locus of many important life and calendar events that bring people together. Annual public yam counting is done in the village. Christmas is often celebrated at a village level. Churches are built in each village, and services are usually run weekly. Residents of a village are constantly brought together in such ways, reinforcing the village as a socially meaningful unit of organisation and belonging.

Earlier I mentioned that the garden hamlet is the most likely place of the true vernacular (section 3), but I have treated the village as the main unit of the speech community for the purposes of this thesis. While it is ethnographically and theoretically sound to treat garden hamlets as micro-speech communities with strong vernaculars, I did not do so for a couple of practical reasons. Firstly, a non-Papuan outsider visiting garden places is potentially socially sensitive. The recording of information by *markai* 'white people' is seen as a legitimising practice, and has the potential to fan disputes that lie between individuals who contest ownership over tracts of land. Secondly, the garden hamlets are quite small and geographically dispersed, with possibly just one or two families occupying a hamlet. Within the time frame of this project, it would have been impossible to collect and process enough data from various garden hamlets in a timely manner. For more discussion of general challenges of data collection in the Morehead area, see chapter 5.

The geographic range of the Nmbo speech community fans outwards from the central Kerake villages. The villages, from north to south, are Govav, Bevdvn, and Arovwe.[2] Earlier work by Rueck (2006) and Tucker, Boevé, Fuller, Gustafs-

[1]There are a handful of individuals based in Daru and Port Moresby, but as far as I am aware they are not part of a larger group of Nmbo speakers. They are not, for the purposes of this study, considered part of the Nmbo speech community.

[2]On official maps with anglicised place names, Govav is known as Gubam, Bevdvn as Bebedebn, and Arovwe as Arufi. Williams (1936) uses the Anglicised names, and Ayres (1983) selectively does. I have chosen to use the forms that better approximate Nmbo pronounciations of the villages.

son, and Rueck (2003) report the three villages as claiming to speak the same language. Govav and Bevdvn are settlements recorded in Williams's ethnography, so have existed as local entities for at least a hundred years. Arovwe is a more recent settlement, which Ayers describes as an area "formerly inhabited by the speakers of *nemnei yem*,[3] who are said to have died out." (1983:140). Arovwe, being a consolidated village, comprises of families that used to live in separate settlements. The families hailing from the now abandoned settlement of Savaram (Tabaram in Williams 1936), just an hour's walk south of modern day Bevdvn, are Kerake and licensed speakers of Nmbo. I will, however, continue to refer to the people of Saravam as *Arovwe hambamn är* 'Arovwe villagers', as this remains consistent with recent work on the language which name Arovwe as a Nmbo speaking village (such as Rueck (2006), Tucker et al. 2003).

I have spoken about Nmbo thus far as though it is a relatively discrete language, but this not the case as demonstrated by the variety variously called Nambo or Namna, spoken by Yarne tribe. It is spoken in the two villages of Drdr (Anglicised: Derideri), and Pongarki (Nmbo: Fongarke). Drdr lies about 6km west of Bevdvn, while Pongarki is another few kilometres west of Drdr. It is a variety that is structurally very similar to Kerake Nmbo, and I will present the linguistic details on a section dedicated to discussing Yarne Nambo/Namna (section 3.4).

Kerake Nmbo and Yarne Nambo/Namna are structurally similar enough to constitute "a common structural base"; one of Labov's criterion for speech community delineation. Ideologically, however, Kerake Nmbo and Yarne Nambo/Namna are identified by their respective words for 'what', which are emblematic of the two tribal groups. They are undeniably part of the same speech community not only because of linguistic structural similarities, but due to the strong ties in marriage that link individual households together. Strictly speaking, it seems that the Kerake village of Govav and the Yarne village of Drdr have the closest ties due to intergenerational reciprocal sister exchange. The close ties were also noted by Williams (1936) during his fieldwork in the 1930s (p.135).

While Govav, Bevdvn, and Arovwe are the central villages of Nmbo, with the Yarne villages of Drdr and Pongarki somewhat on the periphery, there are other villages in the adjoining areas that also constitute part of the Nmbo speech

[3]I am personally unaware of any variety that is called nemnei at this particular point in time. Note that in Ayers' ethnography the Nambu branch languages are described as an actual clause "what is it", so Nmbo is noted as *nambo yem*.

community. This is because of Nmbo speaking Kerake women marrying into these villages. Since it is common for women to marry geographically close to their village of origin, villages adjacent to Kerake locales have sizeable numbers of Nmbo speaking women in their midst. While it is most common for the women to leave their tribal villages, each village has at least one man who has left his tribal village and relocated to his wife's village.[4] Nmbo speaking Kerake women who have left their village of origin will speak Nmbo to their children, and also to other Nmbo speakers in their village of residence regardless of what language is associated with that locale. Married women will continue to use Nmbo when they travel back home to visit their parents, sometimes with husbands in tow; in other words she will have opportunities to continue using Nmbo on a semi-regular basis with her immediate kin.

With this expansive definition, I include villages adjacent to the three Kerake villages as peripheral to the speech community, in the sense that they are geographically peripheral to the central villages. So the peripheral speech community unambiguously involves the immediately adjacent villages of Drdr (Nambo/Namna), Keru (Neme), and Bimadbn (Nen). Based on my data on kinship ties, there are Nmbo speaking women who reside in Fongarke (Nambo), Mata and Morehead (Nama), Mitare (Neme), and Tais (Nä), but these places are further out than a days travel by foot, and constitute the very edges of the Nmbo speech community.[5] A map drawn by Kawas, a Bevdvn man in his late thirties, visually represents these places of interest including a few villages further south (figure 3.1). A schematic representation is also shown in figure 3.2.

Given that exogamy is the pattern of marriage in the area, it is also true that women from non-Nmbo speaking villages marry in to the central Nmbo villages.

[4]The man often leaves his own tribal village due to disputes with other male kin, or because a man must relocate to his wife's village due to debts in sister exchange. Direct sister exchange does not always occur because it is contingent on the availability of an appropriate woman who fits the criteria. It is quite common for a man to find a classificatory sister from among his kin in order to exchange. He is then obligated to return a girl back to the kin member, such as a daughter born at a later date. A man in such a predicament will often live with his wife in her village, until he is able to repay this debt. In each of the Nmbo villages, and in the Nen village of Bimadbn, there was at least one household who was in this situation.

[5]Outside the Yam language area, there are a handful of women from the village of Sigabadru (Täme), the towns of Daru (Kiwai, Tok Pisin) and Balimo (Suki), and also in the capital city of Port Moresby. These non-Morehead area places of residence are not considered part of the Nmbo speech community mainly because the number of Nmbo speakers in each of these places is usually one speaker, and they do not have a significant impact on the rest of the Nmbo speakers on a regular basis. I use the term peripheral purely in the sense of physical periphery, since Nmbo speaking women in adjoining villages are not socially peripheral.

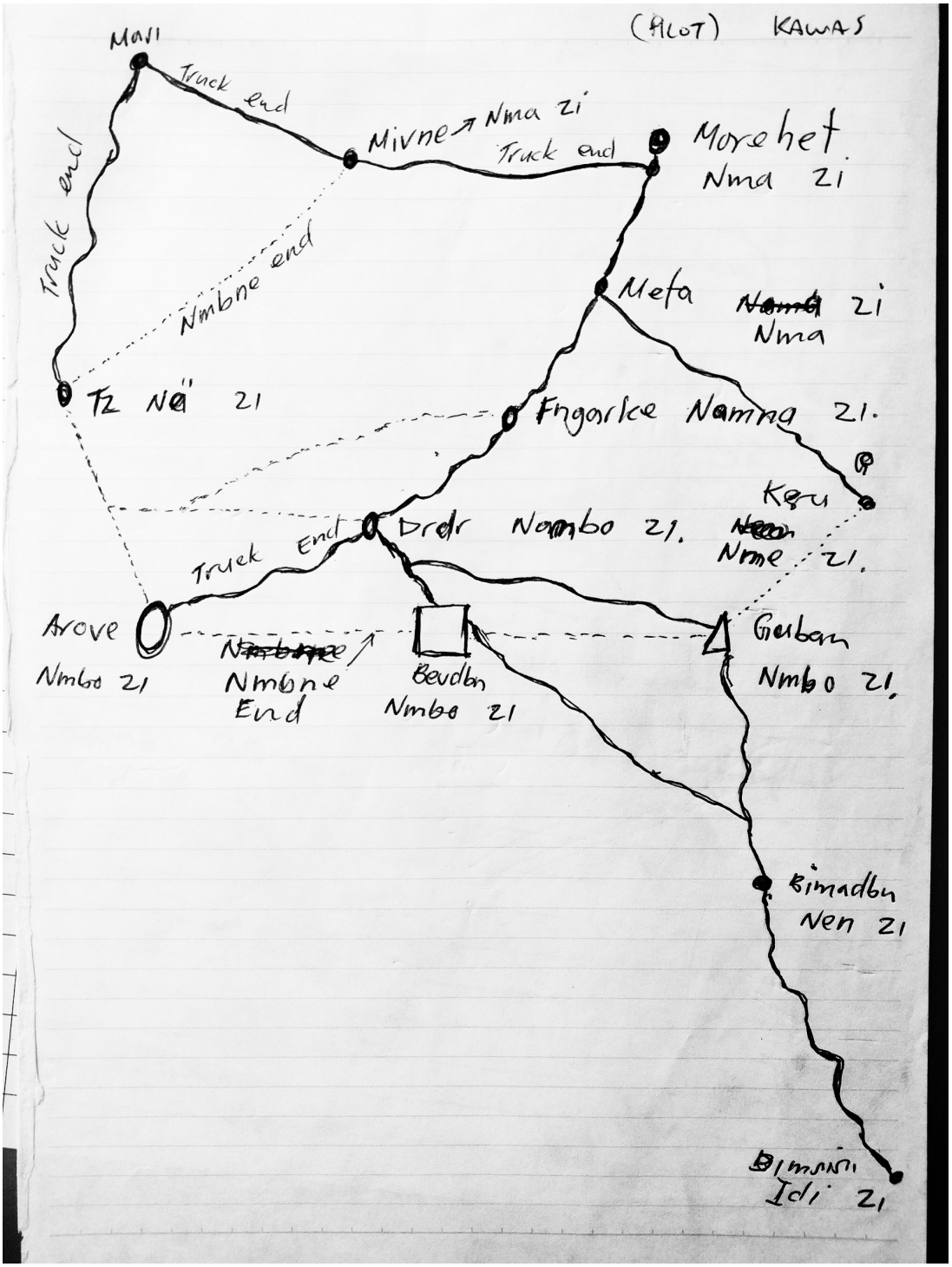

Figure 3.1: A map by Kawas showing the villages and languages around the Nmbo speaking areas. The map is oriented with west at the top of the page.

These members of the village community are also unambiguously part of the speech community. Women who marry in to their husband's village are rarely peripheral to social life. They are encouraged to acquire the language of their husband, since when they do, they may deeply embed themselves in the social life of the village. These women who successfully acquire their husband's language are lauded for their linguistic achievement. On the other hand there does not seem to be, any overt social punishment for not acquiring the village variety. In the villages that I worked with there were a handful of women who expressed discomfort over speaking Nmbo, and some are known outright as not speaking Nmbo at all. These women are, as far as I can tell, not ostracised publicly, and appear to participate in the social life of the village unhindered due to the high levels of receptive multilingualism. In their early typology on the sociolinguistics of exogamy, Stanford and Pan (2013) make a prediction that exogamous marriage patterns of closely related linguistic and cultural varieties are likely to have in-marrying women conform to sociolinguistic norms. As we saw in section 2.7, Nmbo speakers report frankly on their acquisition of maternal varieties, and there were many reports of how Nmbo women acquire their husbands's varieties. In contrast to what Stanford and Pan found with Sui, therefore, the maintenance of their parental varieties by in-marrying women is not seen as in opposition with their learning of their husband's variety: in-marrying women often acquire their husband's varieties, but they also stay loyal to their tribal variety and continue to speak it.

We have discussed the generally high levels of multilingualism in the Morehead area earlier (section2.7), but what this means in terms of language use in individual communities is quite intricate. An important distinction and group to point out are what I call *daily bilinguals*, an imperfect term to indicate the individuals who using two languages daily. These are the women (and occasionally men) who have moved residence to an adjoining village for marriage. This includes Nmbo speaking women who have moved to a non-Nmbo village, or women who have moved in from non-Nmbo villages to a central Nmbo village. These individuals will speak their village variety to their spouse and children. In contrast to these daily bilinguals, the licensed Nmbo speakers who married another licensed Nmbo speaker has fewer obligations or requirements to speak another language in their daily life.

So a schematic representation of the Nmbo speech community is something like the following (a visual representation is presented in figure 3.2): The speech

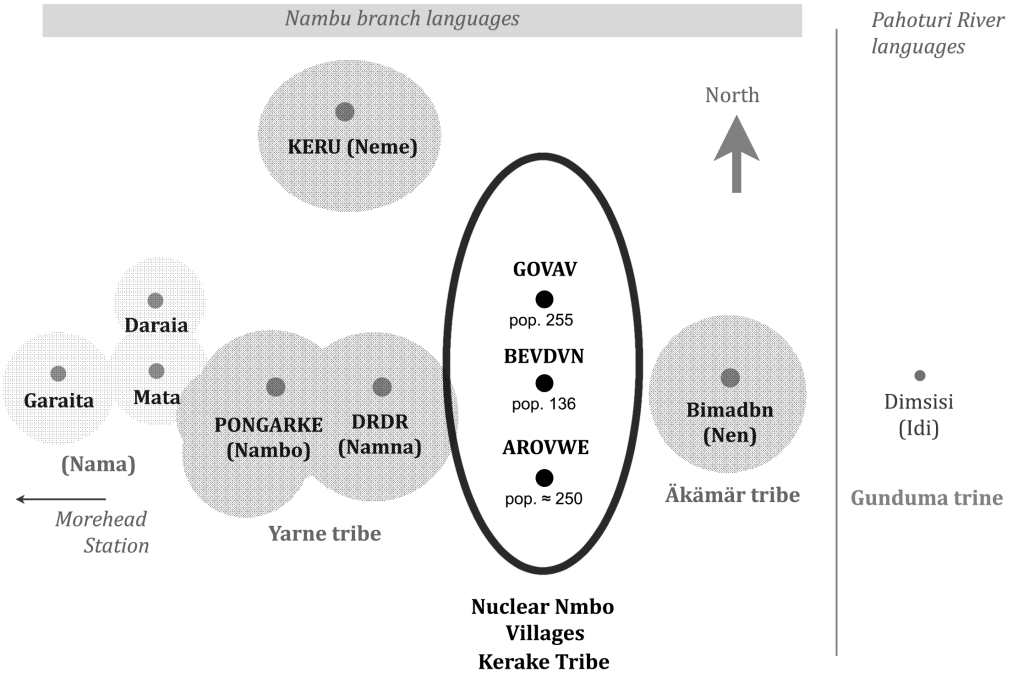

Figure 3.2: Schematic representation of Nmbo speech community. Grey areas represent the geographically peripheral villages that are part of the Nmbo speech community. The Idi speaking village of Dimsisi is represented as it is often mentioned by Nmbo speakers as a part of the Morehead District, but there are no marriage ties to Dimsisi from Govav or Bevdvn.

community of Nmbo has a geographic center and periphery, with the central villages of Govav, Bevdvn, and Arovwe constituting the places of Kerake Nmbo speakers. The closely related Yarne Nambo/Namna variety is part of the Nmbo speech community both in terms of its structural similarity with Kerake Nmbo, and in terms of the social connections through marriage. Marriage networks extend the Nmbo speech community out to non-Nmbo villages on the geographic periphery. The villages of Bimadbn (Nen), and Mata (Nama), are of particular significance to this thesis. The members of the speech community are all highly multilingual, but the in- and out-marrying women are characterised as daily bilinguals who use two or more languages more frequently than Nmbo speakers who are married to another Nmbo speaker. With this in mind, we will now move on to some basic demographic facts about the Nmbo speech community.

3.2 Village Demographics

The total population of the nuclear Nmbo villages in the years 2014-2017 was roughly 710. Govav and Arovwe are about the same size with approximately

250-300 residents, and Bevdvn is clearly the smallest with 100 or so residents. If we include the number of Nmbo women who have married out of the nuclear Nmbo villages, plus their children and husbands who speak Nmbo, we may have about 1000 speakers of Nmbo in the entire Morehead area. Williams (1936) estimates about 700 to 800 as the total population for the entire Kerake tribe in the 1920s (p.47), but this figure could have included speakers of other Nambu branch languages, i.e speakers of Nama, Nen, Neme etc. My figures then show a slight increase in population from Williams' time, which likely reflects the effect of the brief period of colonial presence which saw the use of Western medicine, and rearrangement of villages to fit standards of sanitation required by the administration (Ayres 1983:14-16). Williams feared at the time of his writing that the population of the Morehead area was much diminished due to an outbreak of influenza that spread through the area in 1919 (1936:47-48), but the Kerake population today is well and vital.

Bevdvn and Govav are the two villages I spent the most time in, and of which I have the most detailed knowledge. I have more of an inherent familiarity with Bevdvn, since that was my host village of residence for the duration of fieldwork. Due to the physical and logistic constraints of working across different locales, the village of Arovwe is ill-represented in my work. I worked closely with particular residents of Arovwe, and while I visited the village a few times and have some basic understanding of the lay of the land, my knowledge is significantly impoverished compared to what I can say about the other two villages. With that caveat, I will proceed to explain the lay of the Nmbo speech community.

Bevdvn [beβ.dɐ.βɐn] is a village of 136 individuals comprising 20 households in the year 2015. The people who originate from this village are all descendants of two brothers. As such is it an unusual village where heads of houses are all from the same section, the Sangara. It is geographically located between Govav to the North and Arovwe to the South.

Govav is a much larger village with greater demographic diversity. There were 255 individuals comprising 42 households in the year 2015, and three sections (Sangara, Bangu, and Mayawa) are represented in the village (though the Sangara are represented only by one male lineage and is much diminished). There is technically a community school in the village, but it was operating only sporadically during the course of my fieldwork.

Arovwe is the southern-most Nmbo village, with a population close to that of

Govav at over 200 people. Bangu and Sangara sections are both represented. Arovwe is settled on the banks of the Massi Kussa River with a slightly different micro-ecology to that of Govav ad Bevdvn. The village was the residence of the Australian missionary Graham Martin during the 1970s, and has an elementary and primary school which were both in operation during my fieldwork.

While Bevdvn and Govav villages are only an hour's walking distance apart, their linguistic-demographic configurations are quite different. Govav is the larger village in terms of population size, and is heterogenous in terms of its section composition. An additional important difference regards the levels of multilingualism in each village. I collected data on household couples and their affiliate language in order to get a measure of how multilingual a given village is. For example if a household consisted of a husband from Bevdvn and a wife from Mata, this is counted as a Nmbo-Nama bilingual household. Figure 3.3 shows the proportion of households in Govav and Bevdvn respectively, with Govav showing greater multilingualism. About a quarter of the households in Govav have in-marrying Neme women, and about a fifth are Nen speakers. Another fifth are speakers of Yarne Nambo/Namna.

These figures form my impression of the different degrees of multilingualism across the Nmbo villages, namely that Bevdvn is slightly less multilingual compared to Govav. This is not to say that Bevdvn villagers are not multilingual. Interviews with Bevdvn villagers show individuals have high degrees of multilingualism due to their unique biographies and extended kin. Bevdvn merely has higher numbers of Nmbo-Nmbo households where the daily language is undoubtedly heavily skewed towards the usage of the village vernacular.

The characteristic of Bevdvn as less multilingual is partially explained by Bevdvn's geographic position, flanked by two other nuclear Nmbo villages in the north and south, and having Drdr in the west which is shifting towards Kerake Nmbo as the majority language of the village. Govav on the other hand is flanked by Bevdvn in the south, but has the Nama speaking Garaita directly west. The Neme speaking village of Keru far in the north-west, and Nen-speaking Bimadbn not too far south-east. The Yarne village of Pongarki is accessible by a dirt road. Given the tendency for women to marry into adjoining villages, Bevdvn ends up with a higher proportion of in-marrying women who are themselves licensed Nmbo speakers, while Govav has more options in its immediate area. Govav thus ends up with a much higher proportion of multilingual households, giving the overall

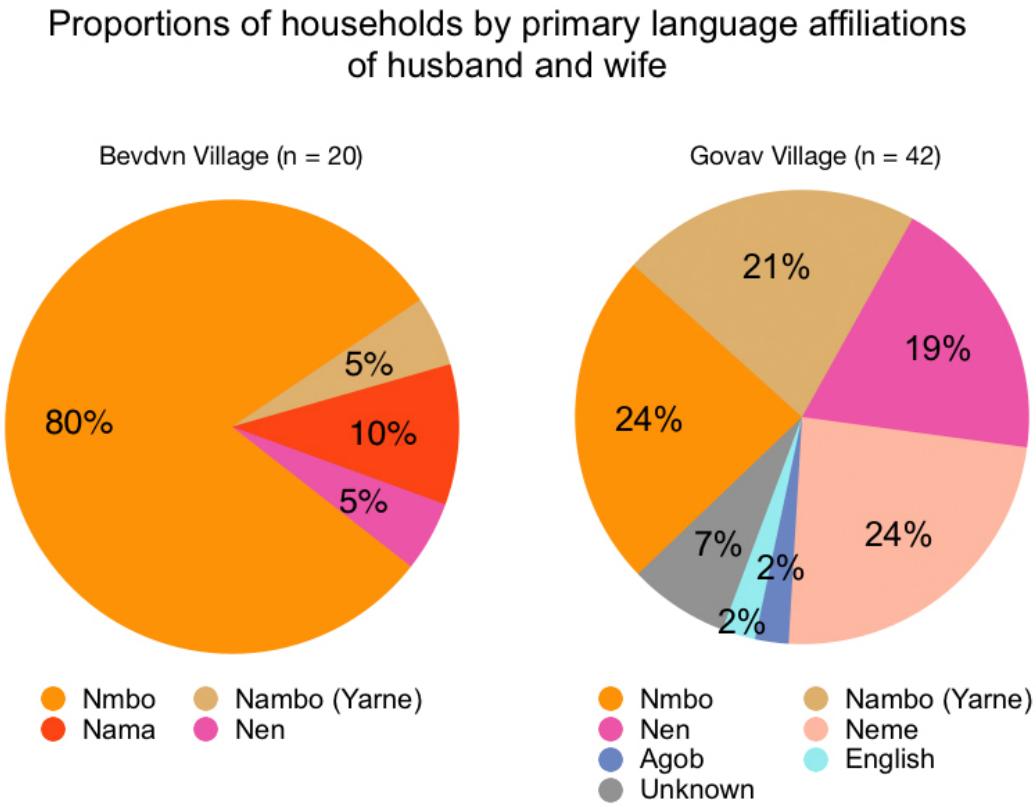

Figure 3.3: Proportions of households by primary language affiliations of husband and wife. e.g. 80% of households in Bevdvn have both spouses speaking Nmbo, while 10% of households have at least one of the spouses a licensed Nama speaker.

village a more multilingual profile.

3.3 Local Social Categories of Relevance

Nmbo speakers are not only sensitive to different languages in their area, but also to differences within the Nmbo language. We will now explore localised manifestations of three cross-linguistically common social variables in order to understand the awareness of intra-language variation: age, gender, and what I call the *Other Village*. I will describe to the best of my knowledge how these three parameters manifest in the Nmbo speech community. Understanding the socio-cultural particulars of age, gender, and the Other Village are necessary in order to analyse the output of the quantitative studies presented later on.

Age and gender are closely intertwined social parameters in the Morehead area, but I have chosen to discuss these separately for two reasons. One is that age and gender express different aspects of social power within the Morehead area,

Figure 3.4: Images of Kerake villages. Top row of Bevdvn, top-right the phone tower is visible.
Bottom-left the village of Govav. Bottom-right an areal shot of Arovwe.

and internally within the Nmbo speech community. Age is undeniably the most significant and cross-cuts gender, with both men and women accorded respect due to their seniority. The other reason for discussing age and gender separately is in recognition of how the combination of the two parameters create distinct social categories with differing associations and communities of practice. The social categories are, broadly speaking, senior men and senior women on the one hand, with young men and young women on the other. Middle aged men and women are their own social category with a set of rights and responsibilities, and communities of practice. We will begin by exploring the parameter of age, and then gender, but with the understanding that they in fact cross-cut each other.

3.3.1 Age

Age and seniority are the most important parameters which inform local concepts of language correctness. The notion that older speakers speak more correctly comes from the wider world view on who has the authority to speak about the world. According to Ayres (1983), it is senior men who hold the rights to esoteric knowledge of places. Senior men are the rightful tellers of stories associated to place, and "only he is considered to "*really know how to tell the story properly*" " (my emphases, Ayres 1983:329). In other words, older people are viewed as the holders of culture and knowledge, and are accordingly given respect and authority by the community.

The elderly, and elderly men in particular, command the most socio-political power in the Morehead area. The implication of older people being more socially and politically powerful is that younger people are less so. As discussed earlier, the effects of modernisation on the Morehead area are quite minimal, so socio-political power structures associated with age and seniority are relatively undisrupted. Some socio-political power has been diminished, such as with the loss of cultural institutions like the male and female initiation cults, but marriages and sister exchanges are still mostly negotiated by older men. A handful of young men and women travel to Daru and Balimo to gain education and training in technical skills, but these processes of modernisation have yet to affect the general respect older men are afforded in the community.

While men aged in their 40s and 50s have the strongest command of English, and therefore social power by access to communities outside the confines of the Morehead area, these middle aged men often defer authority to the elderly men to make decisions on behalf of the community. An anecdotal story told to me demonstrates the finality with which older men can affect a decision. The village of Bevdvn was apparently in negotiations with the company Digicel who had placed a phone tower in the village. The content of the negotiation pertained to the villagers of Bevdvn receiving a small payment for the maintenance of the tower. A middle aged man, at the time a young man, was nominated as the recipient of the salary because of his command in English, and because he had attended high school in Daru. An old man shouted down the decision, claiming that he should receive the money because he was one of the eldest in the village. No one in Bevdvn challenged him, but they also did not facilitate any deal between Digicel and the old man. Eventually Digicel gave up, and no one in

Bevdvn received a salary. I have not verified this story with anyone outside the community, but it is illustrative of two things. Firstly, it shows how an older man may feel entitled to have his opinion trump a collective decision. Secondly, it shows how younger men do not challenge this. In my experience, young men aged in their early thirties to twenties were never involved in the political decisions of the village as a whole. The young men and women of the Morehead area and the Nmbo speech community show great respect to their elders, and the advent of some aspects of modernity have not changed this.

The locus of correctness, both linguistic and cultural, lies in the past. The speech of older people is considered the true and correct way of speaking Nmbo. Nmbo speakers will explicitly state that older people speak true or correct Nmbo, and that I should ask older people about the language in order to document it accurately. I take this belief in older people speaking a true form of Nmbo to mean, on some level, that a Nmbo speaker should speak like older people. The idea that people in the past spoke a language better is a cross-linguistically well observed one, although speech communities tend to express this sentiment more negatively by saying young people speak poorly (e.g. Tagliamonte and Denis (2008) and the notion of *linguistic ruin*). The excerpt below is from an elderly Bevdvn woman who had married to Pongarki village many decades ago. Note that while she judges young people for speaking incorrectly, she herself uses the Namna version of the third non-singular copula in the remote past *tamorwn* (in Nmbo it would be *tamaro*):

> *Dändäi devave gs tamorwn, ymovem huraero zi taitotawt. Nmbo zi. Gnoso gs ym... totr toge ge ym nätambn ziwt. Fonata zi nowavtat gnoso.*
>
> The old people, they spoke only true. The Nmbo language. Now, you see... New children they don't know how to speak. They speak differently today.
>
> WSEK1-B20150813-03PerceptualHZ, 00:04:16.703 - 00:04:26.692

This sentiment of "older people speak true" is shared by younger members of the community. A married man in his late thirties echoes the words of the older woman given above:

> *Dändäwär, zi got dena... [...] hure kt ä nowavta. Gnoso gym dena...totr toge-toge kowavtange... yna zi got, enah. Bä fonata zi got ym.*

> The old people, their words were like... they will speak true there. Today, you know, it's like... young people will speak... these words, sideways. It is a different word/They are different words.
>
> WSEK1-B20150907-04PerceptualKS2LM, 00:13:16.022 -00:13:27.374

The expression of older people speaking more correctly could be a reflection of the belief of the past as better, but I believe it is a common turn of phrase to express some abstract notion of correct language use. Indeed I could not get concrete examples of exactly how it is that older people spoke. Rather, the onus is on younger people who are supposedly spoiling the language. Plenty of concrete examples are given to express this end, such as the use of certain lexical items and their pronounciations. *Okei*, a borrowing of English discourser marker "okay", is a particularly loathed word. To the dismay of many in the Nmbo speech community, however, it is used by everyone including the elderly. Speakers were particularly disappointed when I replayed a recording made by Ayres in the early 1980s, where an elderly resident of Govav (since deceased) was recorded using the dreaded *okei*. Another lexical item associated with younger speakers is a particular use of *bolo* 'old man' or 'poor fellow', but I was unable to fully comprehend how this word was being misused in the eyes of my consultant.

Age and seniority mark the height of socio-political power in the Nmbo speech community, and the Morehead area more generally. Correspondingly, the ideal way of speaking Nmbo is expressed as speaking true Nmbo like the old people. To the best of my knowledge, however, this ideal is not expressed with concrete examples.

A final comment to make about age in the Nmbo speech community pertains to the notion of age cohorts.

3.3.2 Gender

Gender in the Kerake villages is highly socially differentiated, but this differentiation does not readily translate into linguistic differences in the contemporary Nmbo speech community.

Morehead communities are highly patriarchal. SNG communities can be characterised as egalitarian in terms of lacking economic and political stratification, but labour and physical mobility are highly stratified by gender. From a young age, girls are required to help with the chores of the household while her brothers are

not. My neighbours would often send their daughters to go fetch water and tend the fire, while the young girl's brothers of a similar age would be elsewhere in the village. Women bear the brunt of work in daily life. When travelling between villages, men will often travel ahead carrying their hunting gear, while their wives and other women will follow a few meters behind, carrying heavy loads of food, cooking utensils, small children, mats, brooms, and other tools of domestic work. Men and women sit on opposite sides of the church aisle, although this may be an arrangement that was imported with evangelical Christianity. Yet social life is not completely segregated, and women do not publicly express a sense of injustice. At night one will typically find both men and women of a household sitting together around the fire, often with their neighbours both male and female. Young girls and boys will play afternoon sports together. I asked a few women about how they grew up compared to their brothers, and women mostly expressed positive emotions. They explain that there is a division of labour, and each gender does what they are supposed to do; e.g men go hunting, women do the cooking.

> *Tvende herge ge tmaron, ämbru, ynd mer awareñh kawareñtawm. Ẽ, tanzo hakr-hakrava [...] Ynd ägävuwt kovngongayn. Bä däv kowanongaynd, älet.*
>
> Our lives were one. We lived well together. Yes, with my own brothers [...] I would start work, then they would start hunting.
>
> Translated by TT, RM, minor additions by EK. WSEK1-G20170617-01Alqi02MQ, 00:14:01.340 - 00:14:37.889

A few other subsistence activities are worth mentioning to portray the gender stratified nature of work. Fishing and hunting are the primary means of getting animal protein, since animal rearing is only practised in an opportunistic way. [6] Women are the primary fishers, and will go to rivers in groups of anywhere up to ten people. Fishing seems to be done either on a household basis, or as a group of women from neighbouring households. Men will often go out hunting, and this is either done alone or in a larger group of about five people. The hunting expeditions that I know of were of men from the same village, and cross-cut

[6] The Kerake do not practice animal husbandry in the way that is practised in the highlands, with domesticated pigs or poultry. This has also been observed as a cultural tendency over east for Gidra speakers (Akimichi, 1998), and west for the Marind (Olsson pers. comm.). Typically a piglet or young cassowary is captured in the wild, and put in a pen until they grow and are ready for slaughtering. Some individual Nmbo speakers have given me the impression they find pigs in particular unclean or unsettling, but I do not believe this is a widely held impression.

sections. Solo hunts seem to happen at night.

Young unmarried men are given a lot of license to be free, both in terms of physical mobility and social obligation. Unlike young women who are mostly helping their mothers with domestic chores, young men will often be conspicuously absent from the village. Young men tend to hang out with each other, in the sense that they are in each other's company without necessarily any purpose in engaging in set activities. As far as I could tell they hunt, joke, gossip, and in recent years, cultivate and smoke marijuana. They often hang out in the village, but more often than not in the areas around it. There does not seem to be any strict governing of who one should hang out with; only that they are other young men of roughly the same age. There does not appear to be any preference to hang out with young men from the same section, or any other kin affiliation. The most common denominator appears to be that the young men are from the same village, presumably because they happen to be in geographical proximity of one another. Groups of young men will often travel further afield from their village, sometimes for the purpose of attending events in nearby villages such as dances, wakes, feasts, but other times just to go for a wander. There are many stories from men of varying ages telling stories about a time they went to Balimo, or Daru, seemingly just because they could.

The characteristics of language associated with different genders are of a general kind. When asked about differences in language between men and women, some Nmbo speakers will say that they speak the same or in similar ways (example below). On the other hand, when speakers choose to highlight the differences between men and women, they comment on the fact that women marry in from another village, and therefore are speaking another language. Some speakers alluded to differences in pitch between men and women, but it was surprisingly rare to hear this comment.

Men and women speak the same:

> JY: *Dmave dranma navot?*
> HZ: *Bä dena nowavt är dranma nowavt.*
>
> JY: How do women speak?
> HZ: They speak how the men do.
>
> WSEK1-B20150813-03PerceptualHZ 00:04:47.669 - 00:04:49.959

They speak differently:

Gs em zi km mure fnatae ämb nngowavtat, tvende dmave. Ynan gs ym ämb taiman yndvem ah- ämb evhon tamndongm dena "yao yna zi nowavtam ynd, mämä ym e dena knowavta". Kt ädi evrendam yndvem yna, yndon tanzo dmave yvrendan. "Yao yna zi wa nowavta däv."

Our wives utter their voices/accents somewhat differently. Sometimes, ah, sometimes we tell them, "we don't speak that language, it's bad to speak like that". We correct them there and then, I often correct my wife. "You must never speak that language again."

Translation by LM and KS2, major additions by EK. WSEK1-B20150908-04PerceptualKS, 00:12:01.551 - 00:12:16.817

The linguistic differences across gender were possibly more salient in the past, but the differences were limited to the use of select lexical items. Before the increase in Christian missionary work, adult Nmbo speakers had secret words of spiritual power and mythical knowledge as part of their lexicon. Williams and Ayres both record secret words used by men, where common Nmbo lexical items had other secret names that were used only in the presence of other men who had the rights to hear these words (Williams 1936: 72, Ayres 1983:332-333). During the course of my fieldwork both male and female consultants would say that they no longer kept secret words hidden because they are now Christian and do not hold these old beliefs. Indeed, I was given the impression that some secret words were no longer being transmitted to the next generation. Older people still know some of these words and were keen to leave records of it. I had the privilege of recording two elderly women describe and utter secret women's words, but I have been unable to transcribe the recordings for this thesis project. One of my consultants, a woman in her early fifties, described hearing the words at the recording sessions as a thrilling experience. She had told me, "My heart was beating fast. I had never heard these words before, I was never taught them."

The traditional possession of secret words show that women possessed social power within their own spheres. Some of the power that women traditionally wielded has eroded due to the stamping out of traditional rituals and customs, like the use of secret words. Some of my female consultants told me the secret words were dangerous to men, and women would sometimes threaten men with the use of these words if they were angry or felt in danger. Women also took part

in initiation rituals of their own, where older women will lead the initiation and make demands of behaviour from the younger women around the consumption of food. As in most human societies, age and gender cross-cut and interact in various ways, and this is no exception in the Morehead Area. Age and seniority is a major determinant of social power, while other institutions such as the church are not clearly so.

As we will see in the quantitative chapters, contemporary linguistic differentiation of gender is not major in the Nmbo speech community for the variables selected. This is despite the clear demarcation of gender in terms of non-linguistic practice, and the historical practices of selective lexical differentiation. The reason for the non-significance of gender in the Nmbo speech community is likely because other languages and places are more important social entities against which cultural identity is oriented. In the following subsection we will see how the Other Village is spoken about as a place of difference.

3.3.3 The Other Village

Place and language are important markers of identity that people in the Morehead area orient against in the process of their own self definition. The highest level of social organisation that Nmbo speakers distinguish themselves from are, therefore, speakers of other languages and places. But this tendency to differentiate against people of different places is also manifested *within* the speech community of Nmbo. Namely, Nmbo speakers orient against other Nmbo speakers from different villages.

Villagers of Govav orient against villagers of Bevdvn and Arovwe, while villagers from Bevdvn and Arovwe orient against Govav villagers. Bevdvn and Arovwe are treated as a single group according to my consultants. Bevdvn villagers will report that Arovwe villagers "speak the same" as them, and the assessment is returned in kind by Arovwe villagers. The assessment of Bevdvn and Arovwe speaking the same seems based on the history and kin relations tying the people of the two villages. Earlier I mentioned that Arovwe is described as a new village formed during the colonial administration (section 3.1), and the Arovwe villagers I worked with identify as people of Savaram. Bevdvn village is a single section village with descendants from two brothers, and these two brothers had relations with the people of Savaram. When I tried interrogating linguistic differences between Arovwe and Bevdvn, this was taken to mean differences between Savaram

and Bevdvn. Ultimately, there is no difference reported between Bevdvn and Arovwe/Savaram. I take these self reports as true for the sake of this thesis project, and I will refer to Bevdvn alone in the following sections for the sake of brevity. This does, however, include the people from Arovwe and Savaram.

The villagers of nuclear Nmbo villages assess and report the residents of the Other Village as producing a slightly different variety of Nmbo to themselves. One of the first assessments I heard according to Bevdvn villagers was that Govav speakers speak *krtae*, 'heavily' (*krta-e*, heavy-DAT), while they themselves speak *ofae*, 'lightly' (*ofa-e*, light-DAT). Govav villagers reciprocate this assessment by describing Bevdvn and Arovwe villagers as speaking *ofae*, 'lightly', and themselves as *krtae* 'heavily'. The translations of *ofa* and *krta* as 'light' and 'heavy' reflect the English translations my consultants offered. Nmbo speakers will construe these words as positive or negative depending on their perspective. I have heard Bevdvn villagers extrapolate *ofa* as being *mngär* 'quick' (possibly connoting competence), and *krta* as what I understood as 'clumsy'. Govav speakers will construe *krta* as *muye* 'strength' and substance.

Nmbo speaker will also use *ofa* and *krta* to describe other languages, but the translations offered are slightly different. When describing other languages, *ofa* can be translated as 'easy' (example below), while *krta* is translated as 'difficult'. The metaphorical extension of light and heavy to mean easy and difficult is akin to the polysemy in German of *schwer* 'heavy' as 'difficult' and *leicht* 'light' as 'easy'. Nmbo has additional words *avrat* 'soft' to mean 'easy', and *tekr* 'hard' to mean difficult, and Nmbo speakers will use these to describe their competence of other languages, e.g. *avrat zi got nowavtan*, "I speak the easy words". As far as I am aware, Nmbo speakers do not use *avrat* and *tekr* to characterise the speech of the Other Village.

> *(Tz zi) bä näyäretan... mure, kamna yngm mna* ***ofa****...*
>
> [The Tz language] I hear it... the easy ones, those [words] that are easy...
>
> (Translation TT, RM with EK. WSEK1-G20170617-01Alqi02MQ, 00:17:59.180 - 00:18:04.174)

Another common expression Bevdvn villagers use when describing Govav villagers is that they 'drag' (INF: *rohah*) or 'pull' (INF: *lmäneh*) their words. These

are quite unambiguously negative assessments.

> Govavmn är ämb taiman ge noavtat dena... **Rohai** zi bä **yrohat**. **Krtae** ä noavtat.
>
> Sometimes people from Govav speak like... They will **drag** the language. They speak **heavily**. (Translation by JY with EK. WSEK1-B20150804-02PerceptualATJY, 00:05:52.533 - 00:05:52.533)

The use of descriptors such as 'drag' and 'pull', along with negative construals of 'light' and 'heavy', are critical in intent. Bevdvn villagers commonly complained that Govav villagers are "spoiling" or "ruining" their language. The relationship between Bevdvn and Govav is a complex one, with public relations being relatively cordial, but with individuals holding grievances against individuals in the Other Village. The following episode (with a Bevdvn-centric perspective) occurred during my fieldwork, and illustrates how tensions between villagers of Bevdvn and Govav can emerge periodically. There was gossip that a young woman from Govav ran away to the bush with a young man from Bevdvn. The woman was already engaged to another man from another village, so her brothers were angry that the man from Bevdvn had seduced/stolen her away. Three young men armed with bows and arrows arrived at Bevdvn one afternoon, and threatened violence against the young man (who was conveniently absent at the time of this visitation). The young men from Govav cut down a papaya tree that belonged to the young man's mother, and set her yam house on fire. None of the young men of Bevdvn met the Govav men with violence, although they approached the scene of confrontation with machetes in hand. A few middle aged men of Bevdvn chastised the young men from Govav and told them to leave.[7] As described earlier, New Guinean communities are noted to be wary of their neighbours (section 2.4). My understanding of the relationships between Bevdvn and Govav is similar in kind; not hostile, generally cordial, but wary. Each view the other as Kerake and Nmbo speakers, but this does not exclude the Other Village from critique or suspicion.

When giving details about the differences between the village varieties, there are a number of shibboleth words that Bevdvn villagers ascribe to Govav speakers. The word for 'cassowary' shows a village difference, where Bevdvn villagers will say

[7]While all of this was happening, most Bevdvn villagers refused to get involved in the dispute. A consultant from Govav was also present at the time, but he also did not get involved. My Bevdvn consultants expressed the "himself" sentiment (Williams 1936: 246, section 2.4) by stating that the dispute was none of their business

äuya [æw.jɑ], while the Govav villagers will use *gukwa*. Interestingly, the Govav word appears to be a loan word from the Suki (Trans-New Guinea) language up north. The dialect difference for the word cassowary is near-categorically adhered to by the villagers of Bevdvn on the one hand, and Govav on the other. In the entire Nmbo Sociolinguistic Corpus there is only one case where a speaker used the form from the Other Village; a man in his mid-thirties from Govav used the word *äuya*, and realised his mistake as soon as he made the utterance.

Other lexical differences show a one-phone difference, often with the Bevdvn variety realising a voiceless consonant, while the Govav version realises a voiced consonant. Some specific examples are 'house post' *sräzr* vs *zräzr*, 'spleen' *qét* [k͡pət] vs *qd* [k͡pɐd]. Other differences are lexically specific, for example the word *svälnḡ* 'fishing net'. When Bevdvn villagers caricatured themselves, they pronounce the final co-articulate consonant [g͡b] sound somewhat devoiced, resulting in a trisyllabic [sə.βæ.lɐŋg̥͡b]. Bevdvn villagers caricature Govav speakers with a more overtly voiced final [g͡b], and follow it with a ghost vowel resulting in a quadro-syllabic form akin to [sɐ.βæ.lɐŋ.g͡bɐ̥].

Morpho-lexical differences are also given as examples of linguistic difference across Bevdvn and Govav. The interrogative pronoun 'how' in Nmbo in given variously as *drandmae*, *drandae*, and *dranmae*. The word is formed with the root *dr* 'what happened' affixed with the demonstrative ablative suffix *-nma/ndma*, followed by the dative/instrumental *-e* (section 4.4.4.9). Bevdvn villagers state that *drandmae* is the Bevdvn form, while they attribute *drandae* to Govav. *Drandmae* was raised as an example of difference without prompting, while similar words such as *yna-nmae ~yna-ndmae* 'from here' and *ktanmae ~ktandmae* 'from there' were only raised as different when specifically asked. I am unsure whether Govav speakers would agree with these attributions. While I have not quantified or systematically checked who is using which forms, my impression is that all three forms are used by speakers from all three villages.

Kerake speakers of Nmbo are clearly sensitive to variations within their own speech community. Dialect differences in the sense of inter-village differences are explained using similar kinds of language used to describe other Nambu branch languages in the area, such as 'light' and 'heavy', but the nuances are different. Nmbo speakers will give specific lexical items as examples of linguistic differences across the varieties. The Yarne variety of Nambo/Namna is not seen as part of the Nmbo speech community in the eyes of Kerake Nmbo speakers, and does

not factor into Nmbo internal orientations. The importance of place in socially categorising people occurs at the macro areal level, and at the local Nmbo level. At the macro level, people of different places are speakers of other languages. At the local level, people of different Kerake villages are the ones seen as ruining Nmbo.

3.4 The Yarne Variety

Here I will dedicate some space to discuss the Yarne variety of Nambo/Namna. For the purposes of this study, the Yarne villages and their variety are reluctantly put aside to prioritise our understanding of the Kerake Nmbo speech community. The complexities of the situation concerning the Yarne variety could carry a whole study, which is outside the purview of this thesis. As mentioned earlier, the Yarne variety is structurally very similar to Kerake Nmbo, and is more like a dialect of Nmbo rather than a distinct language. The Yarne variety demonstrates a classic instance of how difficult it is to tidily demarcate the boundaries between languages from the linguist's point of view. The emic distinction between Kerake Nmbo and Yarne Nambo/Namna is also an interesting indication of perceived boundaries of a speech community. I will describe what I know about the language and its community here as an instance of sociolinguistic documentation. Additionally, the details presented here serve as a reminder of the constructed nature of the speech community as a unit for academic enquiry; the details of real language communities are socially complex.

The Yarne people are the people of Drdr and Pongarki village, the former only two hours walk from the village of Bevdvn. They speak a lect originally called Nambo, but this language is known as Namna in the eastern village of Drdr. According to the residents of both Drdr and Pongarki, Namna and Nambo are recognised as *ämbro zi* 'one language' (*ämbru zi* in Kerake Nmbo). The reason for the two different names is because of in-law name avoidance. As mentioned in chapter 2 (section 2.3), the peoples of the Morehead area avoid uttering the given name of their in-laws, and will instead use specific kin terms as address and referent terms. The local story tells of how Drdr villagers found themselves in a conundrum when an in-law-to-be had a name that was the same as their language name at the time, Nambo:

> (Note the use of *sens* in the Nambo/Namna transcription, based on the English *change*.)

Zi gi ym, ah... ygärungoynd mnamn... ausa Aaruwav ynadmae... Mätamn är tmorwn Old Man ytqn... Nambo. Okei ynmamn gym, ymovem zi sens yavrotaynd. Ah, ygärngoynd.

The language, they [the Drdr villagers of the past] changed it because Old Lady Saruwav [the Drdr woman who was married] went from here to the old man from Mäta [her husband, the in-law-to-be] his name was... Nambo. OK because of this, they changed the language. Ah, they changed it [lit. They turned it over/around].

WSEK1-Y20151010-13YarneZiYtqnYP, 00:00:24.236 - 00:00:42.051

Ygärngoynd mnat... kamat dvnan, wamsa ytqn tvende yuvi ym dené, kamat yao bi ym ymdetam mrz emo giya tägronga. Kamat ge tmorwn, Nambo, okei [...] Ytqn ygärungoynd. Namna. Okei gnoso ge ym, Namne zi tortr ärm yus yavrotat but, wli-wli ge tamorwn, yna är teikr gsve wavtan. Yvende zigot bi ym, Nambo. Yarnavende zi bi ym.

They changed it because... Because of our *tambu* [in-law], we avoid his name because it's taboo. We cannot call the name of the son-in-law who gets married to a sister, Nambo was their *kamat* [affine] [...] They changed the name to Namna. OK today young people use [the name Namna] but really old people, the old people I've already mentioned, their language is Nambo. This is the Yarne language.

(WSEK1-Y20151010-13YarneZiYtqnYP, 00:00:46.300 - 00:01:31.366)

Thus Yarne Nambo came to be identified by two different names: Nambo in Pongarki, and Namna in Drdr. Since language names in this area are based on the word for 'what', the Drdr village word for 'what' became *namna*. Linguistically this is not as great a stretch as one may imagine, since the word *namna* appears to be morphologically built up from the root *na* 'what reason'. One possibility is *na* taking an originative affix *=mna* (*=mn* in Kerake Nmbo, section 4.4.4.15). Another possibility is that *na* is compounded with the manner demonstrative *mna* (section 4.2.1.3). Either way, the semantics of inquiry are still maintained, and the usage of a new interrogative may be rendered relatively straight-forwardly.

The similarity of the names between the Kerake and Yarne varieties is indicative of other lexical and structural similarities. The two languages are also identified by speakers as being very similar, with one Yarne man describing Nambo/Namna

as "more like a dialect" of Nmbo, and the differences are that of "melody" rather than that of content. A comparison of basic vocabulary in the Yam Finder database (Carroll et al. 2019) finds only nine non-cognate words (table 3.1), meaning the rate of shared vocabulary between Nmbo and Nambo/Namna is 97.6% (373 cognates out of 382)[8]. For comparison, Nmbo and Nen share 83% of the word list, while Nmbo and Nama only share 69%[9]. Lexico-statistical work and network cluster analysis by Rueck (2006) also shows that Kerake Nmbo and Yarne Nambo are very closely related varieties. On the structural side, the independent pronominal paradigm is identical across the two varieties. The paradigmatic oppositions expressed in the ambifixing verbs appear to be identical but with slight differences in suffixal forms. The copula paradigm is very similar, although the Yarne people I worked with rejected the primordial paradigm which is present in Kerake Nmbo (table 3.2 for comparison of the remote past form of the copula). The phoneme inventory is very similar, with noticeable variation in the correspondence set [h] for Nmbo being realised as a velar approximant [ɯ].

Lexical Item	Kerake Nmbo	Yarne Nambo/Namna
arrow head	dofa	bile
yam house	zä rngu	weh mngo
bamboo, bow	nämb	g̅g̅n
yam	nne	wem
bat	drfn	bimbinas
pig	kimb	mamwi
river	aragav	ttkave
flying fox	bwinder	qter
sand	hmg̅ar	dzermband

Table 3.1: The nine non-cognates between Kerake Nmbo and Yarne Nambo/Namna as by the Southern New Guinea Word List. The last two cognates for *flying fox* and *sand* may possibly be a cognates.

It is difficult for an outsider linguist to assess how much shift and/or attrition is happening in a linguistic sense. When I administered the Yam Finder word list to two Drdr villagers in their late forties/early fifties, the male speaker quite con-

[8]The full Yam Finder Word List constitutes 386 words. A few dubious data points were omitted from this calculation.

[9]These percentages are based on a total of 326 words, rather than the current 382. These figures are based on Ellison and Kashima (2015)

P/N	Kerake Nmbo	Yarne Nambo/Namna
1sg	qmaro(n)	qmorwn
1du	tnrwm	tnrwn
1pl	tnmaro(n)	tnmorwn
2sg	knmaro(n)	knmorwn
2du	tnrwm	tärwn
2pl	tnmaro(n)	tamorwn
3sg	tmaro(n)	tmorwn
3du	tarown	tärwn
3pl	tamaro(n)	tamorwn

Table 3.2: Conjugation of Yarne Nambo/Namna copula verb (Fieldnotes 2017:27). The Yarne people whom I worked with rejected the primordial form of the copulas, e.g. *wmowrn 1sg.prim

sistently produced what sounded to me were Nmbo words. The female speaker had married in from Pongarki to Drdr, and her pronunciations of Nambo words were exaggerated for ease of hearing. Even so the differences are minor. For example the Nmbo verbal nominaliser -*h* was variously [-(ə)y ~-(ə)ʊ ~-(ə)ɯ]. Some Yarne words are prenasalised word initially, while the cognate in Nmbo is not, e.g. Nambo *ᵐbalao* 'electus parrot' vs Nmbo *balao*. Some monosyllabic Nmbo words would have a copy vowel after the final consonant, rendering it a disyllabic word, e.d. Kerake Nmbo *sɐ͡kp* 'canoe', Yarne *sɐ͡kpə*. These are already very minor differences to begin with. If the villages of Drdr and Pongarki were included as core parts of the speech community, a variationist might describe these lexicophonological differences as instances of sociolinguistically conditioned variation, with village/tribe as the social variable.

The linguistic differences between Kerake Nmbo and Yarne Nambo/Namna are not large, but the social difference is highly important to both communities. The Drdr villagers and Kerake Nmbo speakers are quite aware of the decline in usage of the Yarne variety. The reason for the decline is identified by the following Yarne man as due to the influx of Kerake Nmbo speaking women who married into the village of Drdr The first excerpt is from a Yarne tribe man, while the second excerpt is from a discussion with a Kerake man:

> *Ämb evhon, gym ämb zi bä nngoavtat dena... Nmbo zi, o, zi tärvär evhwetam dena, Nama zi... Ede yände- Yärnavende zi gi ym, kndun de namdai. Qvn-qvn ym gnoso.*

Some times they [the children of Drdr] speak other languages like this, Nmbo language or we mix it a lot with Nama zi... So the Yarne's language is shrinking [lit. is becoming inside]. It's disappearing today [lit. it is in lots of small holes today].

(Translation BA, minor additions by EK. WSEK1-Y20151010-13YarneZiYtqnYP, 00:02:20.985 - 00:02:34.121)

Tvende mrz-mrz o, tvende nmbo ziär mrz-mrz kt ymsaran, ynan gs ym, mndeva eyäretam mnan woi kitong ädi namndet nmbo zi äi gs.

Our sister, or our Nmbo speaking girls are staying there [at Drdr], so we hear them with interest, that's how our Nmbo language is spreading.

(Translation by LM, KS2, minor additions by EK. WSEK1-B20150908-04PerceptualKS,00:11:06.197 - 00:11:12.955)

The data on demographic shifts in the village of Drdr are quite clear. While I was unable to get an accurate count of residents in Drdr, the Tucker, Boevé, Fuller, Gustfasson, and Rueck (2003) figure of 78 Drdr villagers in the year 2003 (pp.6-7) is close to my impressions when I visited in 2015.[10] Of these 78 people, about twelve people would be married women, the remainder being men and children.[11] According to the figures in Rueck (2006), there are at least 10 women who have married in to Drdr from Govav. My data on in-marriage to Govav shows an approximately reciprocal number with 9 women marrying in from Drdr. This suggests that 10 out of 12 adult women in Drdr are Kerake Nmbo speakers (83%). Even if we take a more conservative measure by looking at percentage of Kerake Nmbo speakers by household using the Tucker et al. (2003) figures, 10 out of the 17 household would be comprised of a spouse with a Kerake Nmbo speaker (58.8%). Williams had noted in the 1920s already that Govav and Drdr exchanged sisters regularly (1936:155), suggesting that the increase usage of Kerake Nmbo in Drdr was facilitated by intergenerational reciprocal sister exchange.

[10] I have compared my population number for Govav against those of Tucker et al. (2003) in order to gauge the reliability of their figures for Drdr. Their figures for Govav village are 260 people, which is only 5 more people than my count of 255. I therefore take the figure of 78 Drdr villagers as reliable for estimating the population for the year 2015.

[11] This figure was calculated by figuring out the percentage of people in a village who are married women. Based on my numbers from Govav and Bevdvn, about 15% of an entire village's population are married women. The figure of twelve married women in Drdr is based on this 15% figure.

The sociolinguistics of Yarne Nambo/Namna is a fascinating site where the distinctions between language varieties, and concepts of speech community, are problematised. It is regrettable that I could not explore the dynamics of variation in the village of Drdr, but this short description provides a rare glimpse into a context of indigenous dialect shift, not under pressure from a national language but another local vernacular.

3.5 Discussion and Conclusion

Labov's criteria for delineating a speech community holds well for an etic identification of the Nmbo speech community. It is fortunate that emic concepts of language, place, and people mostly overlap with the articulated criteria. The locally salient distinction between Kerake Nmbo and Yarne Nambo/Namna, however, demonstrates the classic mismatch between locally important socio-linguistic distinctions and the academic definition of a speech community.

The Nmbo speech community is geographically well-defined as a community based in and around the villages of Govav, Bevdvn, and Arovwe. It is socially well-defined as a community of Kerake people and other tribal peoples who speak or understand Nmbo. The emic association of Nmbo with the places of Govav, Bevdvn, and Arovwe, have made this geographic delineation relatively unproblematic. The inclusion of villages adjacent to nuclear Nmbo villages as part of the speech community accounts for the Kerake women who have relocated to their husbands's villages, and continue to speak Nmbo as a language of daily life. The ability to geographically delineate the Nmbo speech community suggests the importance of place and face-to-face social interaction in characterising this community.

The geographic proximity and social connections between the Kerake villages and the Yarne villages means that the villagers of Drdr and Pongarki are naturally part of the Nmbo speech community. Ideologically, however, Kerake Nmbo and Yarne Nambo/Namna are distinct languages. The structural similarities between the two languages provide linguistic reasons for treating the two varieties as dialects of Nmbo. Ideologically, however, Nmbo and Nambo/Namna are distinct languages because of emic associations between language with tribe and place. The local distinction of Kerake Nmbo and Yarne Nambo/Namna demonstrates a concrete example of what we have been more abstractly talking about as the importance of tribe, language group, and place as social constructs to culturally

orient against (section 2.6). The very small linguistic differences between Kerake Nmbo and Yarne Nambo/Namna are heightened and made salient in local discourse, and ideologically form the basis of distinguishing oneself from the other. No matter the structural similarities, the fact that Nambo/Namna, or Nmbo, is the language of another tribe and place is what matters.

The salience of tribal-linguistic differences does not mean, however, that Nmbo speakers do not comment on community internal differences. Nmbo speakers have a sense that men and women have linguistic differences, but the commentary seems to center around a recognition that women tend to marry in from other villages, and therefore speak languages that are not Nmbo. Age differences are also noted, where older speakers are abstractly defined as the most correct speakers of Nmbo. The greatest distinction that Nmbo speakers make, however, is against the Other Village. Nmbo speakers identify lexical, phonological and lexical-morphological differences across Govav and Bevdvn/Arovwe. Here there is no distinction between peoples of another tribe since the Other Villagers are all Kerake; but the distinction of people as coming from another place is still maintained. We will see in the quantitative investigations further on that the Other Village is indeed a significant social factor in explaining patterns of variation, even for linguistic variables that are below the level of consciousness.

Part II

Linguistic Descriptions

Chapter 4

Sketch Grammar

> "The sketch grammar of a language documentation is a bit of everything. Through most stages of the process of documenting the language, it is a preliminary grammar that needs constant revising." *(Mosel 2006:302)*

This section of the thesis presents a sketch grammar of Nmbo. The purpose of this sketch grammar is to present what Mosel (2006) calls a "Language Documentation Sketch Grammar". It is based off eight months of fieldwork (years 2014-2017[1]), and in consultation with published material on sister languages of Nen, Nama, Komnzo, and Ngkolmpu (citations listed a few paragraphs on). The goal of this sketch grammar to aid interpretation of corpora annotation, and to best reflect our current knowledge of this under-described language. Additionally, this section also stands to provide the linguistic context within which variation is to be found, and provides a glimpse into the linguistic richness of Nmbo. The relevant sections for each quantitative study are listed in their respective chapters.

In the introduction to the thesis I gave a short allusion to the complexity of the Nmbo language, but in this preamble I will provide some additional typological facts to set the linguistic scene. Nmbo can be described as an SOV,

[1]More precisely: six weeks in Sept.-Oct. (2014), three months and two weeks in July - Oct. (2015), two weeks in July (2016), and two months and two weeks in June - Aug. (2017).

or AUV[2] language, with optional pronominal expression of core arguments. The alignment system is ergative/absolutive, and the pronouns make a three way person/number distinction along singular and non-singular lines. Clusivity is not a feature. As noted for other languages of Southern New Guinea (Evans, Arka, et al. 2018:647, Döhler 2018:3), Nmbo does not have features which Papuan languages of the Sepik and the highlands are known for; there is no switch-referencing, clause chaining, or tone.

Nominal morphology is entirely suffixing and inflectional. There are no gender or other noun class distinctions. Unlike the verbal morphology, nominal morphology is analytic, with 14 semantic case markers.

The most striking aspect of Nmbo and other Yam languages is undoubtedly its verbal morphology. The verbal morphology is highly synthetic, and the verb complex comprises a maximum of nine slots that distribute argument indexation and TAM information across non-contiguous morphemes; what is referred to as *distributed exponence* (Carroll 2016 for Ngkolmpu, or *multiple exponence* by Caballero and Harris 2012). At its most complex, the verb can express the following slots with the following overall interpretation:

(5) a. k-ng-a-waramo-ta-ng-e-ay-m

2sg.U:β-AND-APP-give-IPFV-PFV.FUT-PFV.DU-PRET-1nsg.A

The two of us should have given it away to you (sg.)

Fieldnotes 2017, Book 1:16

The terminology employed in this sketch grammar follows the works by Evans on Nen (2017b, 2016, 2015a, 2015b, 2012a, 2009), Siegel (accepted, 2017, 2016, 2014a) and Döhler (2018) on Komnzo. Some analytical and terminological conventions include: analysis of verbal types in terms of valency rather than transitivity, labelling verbal indexation in terms of semantic macro-roles *actor* and *undergoer*, homorganic nasal + stop sequences as prenasalised stops, homorganic bilabial + [w] sequences as rounded bilabials. Conventions such as the interlinear glossing to represent circumfixal morphology (see thesis preamble) are also particular to the Yam languages.

[2]Or as Döhler (2018) put it, it is more accurately Actor-Undergoer-Verb order since there is only weak evidence for a subject category (p.7).

A dictionary is being created as part of the documentation endeavour. At the time of this thesis submission, a 2018 draft version contains 2336 words with multiple senses.

The sketch grammar will cover basic aspects of phonology (4.1), word classes (4.2), description of particle forms and functions (4.3), nominal morphology and case (4.4), the verb complex and morphology (4.5), noun phrase structure (4.6), an overview of inflectional TAM (4.7), simple clause construction with argument marking and valency alternations (4.8), and finish with some examples of complex clause structures (4.9).

Throughout the various sections I have noted variable linguistic phenomena, but these have been combined into Appendix A.19 for ease of viewing.

4.1 Phonology

This section will give a selective phonological description of the Nmbo language. A consonant and vowel inventory are provided in sections 4.1.1 and 4.1.2 respectively. There is a brief description on phonotactics (section 4.1.3), as well as on intonation (section 4.1.4) in order to support the use of intonation units for the quantitative studies.

4.1.1 Consonants

Nmbo has 28 consonants (table 4.1). This is somewhat larger than its geographically closest sister language Nen, which has 23 consonants (Evans & Miller, 2016). The larger consonant inventory in Nmbo is due to its retention of an ancestral rounded labial series and of the bilabial fricatives, both absent in Nen (*^{m}b^{w}, *m^{w}, *f^{w}, Evans, Carroll, Döhler, 2017).

Table 4.2 contains examples of consonant occurrence positions in syllables.

4.1.1.1 Stops

Nmbo has three manners of stop consonants: voiced, voiceless, prenasalised. In addition there is a voiced rounded bilabial stop /b^{w}/. This last phoneme will be covered in the rounded bilabials section (4.1.1.2). Stop consonants distinguish five places of articulation: bilabial, alveolar, velar, labial-velar, and rounded bilabial (the final two occur only as voiced stops). All the stop consonants except

	Bilabial	Rounded Bilabial	Alveolar	Palatal	Velar	Labial-Velar	Glottal
Voiceless Stop	p		t		k	k͡p	
Voiced Stop	b	b^w	d		g	g͡b	
Prenasalised Stop	^{m}b		^{n}d		^{ŋ}g	^{ŋ}g͡b	
Nasal	m	m^w	n	ɲ			
Trill			r				
Voiceless Fricative	ɸ	$ɸ^w$	s				h
Voiced Fricative	β	$β^w$	z~dʒ				
Approximant				j		w	
Lateral Approx.			l				

Table 4.1: Nmbo consonant inventory.

the rounded bilabial /b^w/ have prenasalised counterparts (^{m}b, ^{n}d, ^{ŋ}g, ^{ŋ}g͡b). The regular voiced and voiceless stops occur word initially, but the prensasalised stops do not.

The alveolar and velar stops show a voicing contrast across all positions, and occur word initially, finally, and intervocalically.

1. /t/ vs. /d/

   (a) /tər/ ‘fruit bat’ vs. /dər/ ‘how’

   (b) /tɑ.rɐh/ ‘to dig’ vs. /dɑ.rɐh/ ‘to happen’

2. /k/ vs. /g/

   (a) /kor/ ‘again’ vs. gor ‘footprint’

The bilabial stop contrasts are restricted, with an asymmetry between the distributions of the voiced and voiceless incarnations. The voiceless bilabial stop /p/ occurs word initially only in loan words such as *plen* (/plen/, ‘plane’), or personal names *Patra* (/pɑt.rɑ/). It occurs as an initial onset, and word finally, in reduplicated words such as *pit-pit* (/pit.pit/, a type of plant) or *lap-lap* (/lɑp-lɑp/, ‘sarong’).

The voiced stop /b/ occurs much more frequently than its voiceless counterpart, but also has quite a restricted pattern of occurrence. It occurs word initially,

	Word/Syllable Initial	Word /Syllable Final
p	plen ‘plane’	-
b	bæ(3ABS)	-
t	tɐ^{n}d‘hand’	tot ‘long pointed object’
d	deβe ‘father’	k͡pɐd ‘spleen’ (Govav variety)
k	kət ‘there’	ɸərɐk ‘blood’
g	got (‘back’/’bone’)	ɑg ‘coconut’
k͡p	k͡pəmbiβ ‘sago stamping bag’	sək͡p ‘canoe’
g͡b	g͡bɐg͡bɐn ‘bamboo; bow’	jɐng͡b ‘mat’
m^{w}	m^{w}iti ‘exchange cousin’	-
b^{w}	b^{w}e ‘seed; round thing’	-
ɸ^{w}	ɸ^{w}e ‘tape worm’	-
^{m}b	-	gu^{m}b ‘pool’
^{n}d	-	tɐ^{n}d ‘hand’
^{ŋ}g	-	b^{w}era^{ŋ}g ‘fly’
^{ŋ}g͡b	-	jɐ^{ŋ}g͡b ‘bag’
m	men ‘bird’	jɑm ‘event; way’
n	nɑ ‘why’	wən ‘tree’
ɲ	ɲɑmoh ‘killing for dead’	meɲ ‘nose’
r	rɑjɑ ‘incoming tide’	mer ‘good’
ɸ	ɸɑn ’clear place’	?
β	-	zeβ ‘hair; fur’
s	sɐn ‘tooth’	ɸrɑs ’poison root plant’
z	zi ‘word’, ‘story’	monz ‘leech’
h	hɑ^{m}bɑ ‘village’	wingoh (‘to see’)
j	jənd (1ABS)	wæj ‘again’
w	wiwi ‘mango’	wɑw ‘ripe’
l	lemæneh ‘to pull’	m^{w}il ‘nipa fruit’

Table 4.2: Occurrence positions of Nmbo vowels.

and can occur syllable initially in a reduplicated form, e.g. /bɑn.bɑn/ 'under', 'shadow', /bər.bər/ 'fear'. It does not occur word finally, or intervocalically.

Finally Nmbo has labial-velar coarticulated stops /k͡p, g͡b/. Some speakers do not co-articulate the two stops as much as others, so the realisation of the phone is sometimes closer to [k^{w}] or [g^{w}].

3. /k/ vs. /k͡p/
   (a) /kɑ/ 'where') vs. /k͡pɑ/ 'Torresian crow'
   (b) /kɑ.kɑ/ 'near' vs. /k͡pɑ.k͡pɑ/ 'chionanthus ramiflorus'
   (c) /ka.ki/ 'grandparent; grandchild' vs. /k͡pəki/ 'vessel'
4. /g/ vs. /g͡b/
   (a) /gə.gə.n/ 'rock' vs. /g͡bə.g͡bə.n/ ('bow', 'bamboo')
   (b) -/jɐŋ.g/ (perfective remote past 3SG agent suffix) vs. /jɐŋ.g͡b/ 'bag'

Labial-velar stop consonants are cross-linguistically uncommon, but concentrated around certain parts of the world such as West-Africa (Cahill 2000:71), and Vanuatu (Billington, Thieberger, and Fletcher 2018). Within this distribution, Nmbo shows a rare phonemic contrast between voiced and voiceless labial-velar stops. It is in fact so rare that Cahill suggests a phonemic /k͡p/ "may be non-existent" (Cahill 1999:157). This makes Nmbo, and her sister language Nen for which this distinction is also present (Evans and Miller 2016:3), two of the few languages known to make this voicing contrast between labial-velar coarticulates.

5. /k͡p/ vs. /g͡b/
   (a) /k͡pər.k͡pər/ 'bushfire' vs /g͡bər.g͡bər/ plant of species Acronychia
   (b) /də.k͡pən/ 'black anthill' vs /də.g͡bən/ 'joint'

The difference between the plain velar stops and the coarticulated labial-velar stops is also acoustically visible. The regular velar stop shows a stop release burst leading into the following vowel (figure 4.1). The coarticulated stop shows an increase in intensification before leading into the vowel, and the stop release shows a lot less frication (figure 4.2).

All voiced stops except the rounded bilabial have a prenasalised counterpart: /^{m}b, ^{n}d, ^{ŋ}g, ^{ŋ}g͡b/. These occur in all positions except word initial position. A preliminary reconstructions of Proto-Nambu posits word initial pre-nasalised

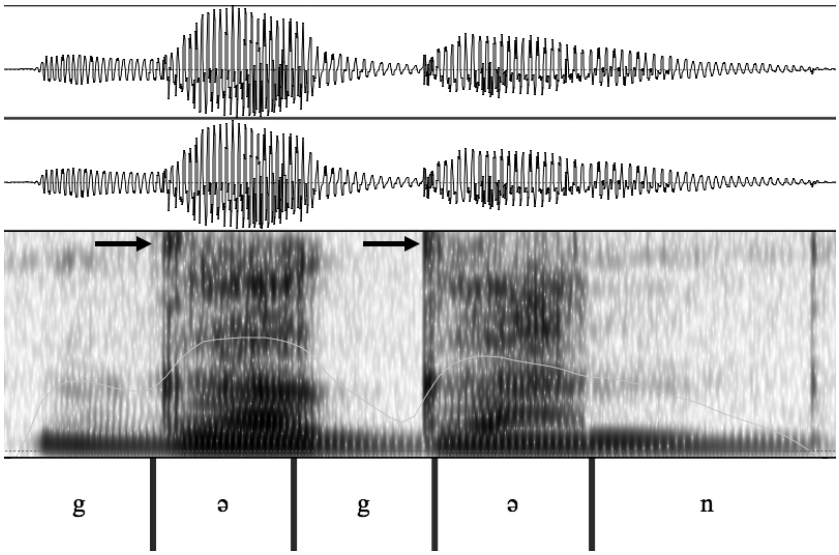

Figure 4.1: Spectogram of [gə.gən] 'rock' as pronounced by a middle aged female speaker from Bevdvn. Arrows indicate the frication found upon release of the velar stop. Contrast to figure 2 below.

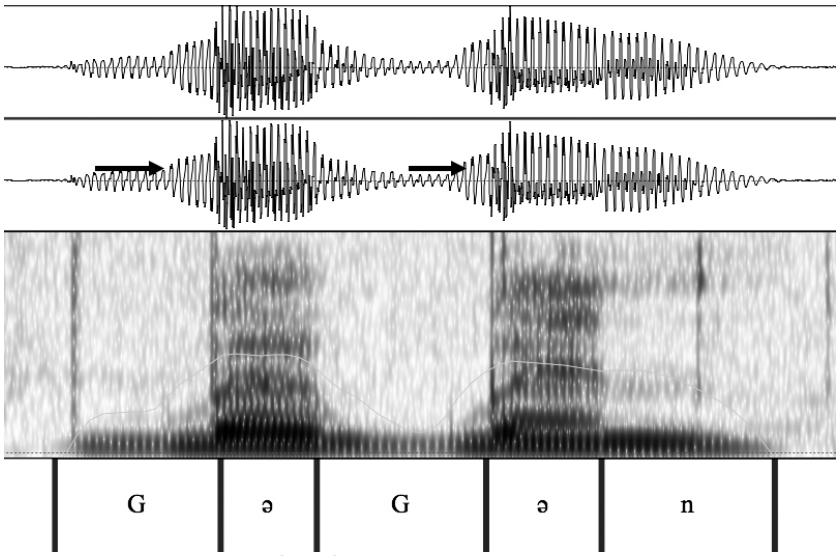

Figure 4.2: Spectogram of [g͡bə.g͡bən] 'bow; bamboo' as pronounced by the same middle aged female speaker from Bevdvn. Arrows indicate an increase in intensity before entering into the vowel. Contrast to figure 1 above. Also note that the release of the velar stop shows significantly less frication than the [g] release shown in figure 1 above.

stops (Evans, Carroll, and Döhler 2017), and Nmbo appears to be one of the many languages in the Nambu branch that have de-prenasalised. At non-word initial syllable boundaries the prenasalised stops syllabify the nasal as the coda of the preceding syllable, and the the stop will form the onset of the following syllable, e.g. *ba^{n}d=an* (land=LOC) > /bɑn.dɑn/. The minimal pairs below contrast the prenasalised stops with their non-prenasalised counterparts in such a fashion.

6. /b/ vs. / ^{m}b/

   (a) /bər.bər/ 'fear' vs /bərm.bər/ 'mouth'

7. /d/ vs. /^{n}d/

   (a) /ku.du/ 'Southern Crowned Pigeon' vs /kun.du/ a type of drum

8. /g/ vs. /^{ŋ}g/

   (a) /ɑg-em/ coconut-ERG vs /ɑŋ.gem/ 'walking up and down repeatedly'

   (b) /i.gɑ.rɑ/ male personal name vs. /iŋ.gɑ.rɑ.iŋ.gɑ.rɑ/ (a type of plant)

4.1.1.2 Rounded Bilabials

Nmbo has four rounded bilabials: /b^{w}/, /m^{w}/, /ɸ^{w}/, and /β^{w}/. They occur word initially and intervocalically, but not word finally. The pronunciations of these consonants vary, especially when followed by [i] or [e]. For example [m^{w}] is sometimes realised with less rounding, e.g [mɑm^{w}i ~mɑmi] 'pig'. The reverse also happens where unrounded stop consonants get labialised by some speakers, e.g. [mi jɐm ~m^{w}i jɐm] 'It is still so', [æ^{m}bæ^{m}be æ^{m}bæ^{m}b^{w}e] 'sometimes'. Some words, such as the village of Arovwe [ɑroβe ~ɑroβ^{w}e] vary so much by speaker, a more systematic study will need to be undertaken to determine what patterns the variation. This is to say that the rounded bilabial consonants are rare phonemes in Nmbo phonology, and exhibit a lot of variation.

Minimal pairs of these phones are rare. So far there is only one minimal pair for /b^{w}/ contrasting with /b/, and /β^{w}/ with /β/. Minimal pairs contrasting / m^{w} / and /m/, /b^{w} / and /ɸ/ have not been found, but there is a contrast between /β^{w}/ and /ɸ^{w}/.

9. / b^{w} / vs. /b/

   (a) /b^{w}e/ 'seed; round thing' vs /be/ 2sg.DAT

10. /βʷ / vs /β/

    (a) /kɑβʷe / ‘tree branch’ vs /kɑβe/ ’cockatoo’

    (b) /bɑβʷɑ/ Singaporean Taro /bɑβa/ sister’s children/mother’s brothers

11. /bʷ/ vs /ɸʷ /

    (a) /bʷe/ ’seed; round thing’ vs /ɸʷe / ‘tape worm’)

The current evidence for postulating /mʷ/ as a phoneme comes from comparison with Nama and Nen (table 4.3). Nama is one of the more phonologically conservative Nambu languages with a retention of pre-nasalised stops, including the prenasalised rounded bilabial /ᵐbʷ/ (Evans et al. 2017). Nmbo, like Nama, retains word initial /mʷ/ from proto-Nambu, which has been lost in Nen to the east. Interestingly there are very few shared cognate words with /mʷ/ between Nama, Nmbo, and Nen; for example, Nmbo *mamʷi* ‘pig’ in both Nama and Nen is an entirely different word, *kiemb*.

	Proto-Nambu	Nama	Nmbo	Nen
’exchange uncle’		mʷitareɸ	mʷitareβ	mitarbe
’exchange aunt’		mʷitartəm	mʷidædem	mitadma
’exchange cousin’	*mʷiti	mʷite	mʷite	miti
‘jaw’	*ɸətkamʷe	ɸət	ɸətkamʷe	bətkam

Table 4.3: Cognates of /mʷ/ words in Nama, Nmbo, and Nen. Proto-Nambu resconstruction from Evans et al. (2017):17

4.1.1.3 Nasals

There are three nasals: /m/, /n/, and /ɲ/. There is no velar nasal, except as the prenasalisation of the voiced velar obstruent /g/, e.g. /ᵑg/. All nasals can occur in all positions. As mentioned in the labialised stop consonant section /m/ is realised by some speakers as [mʷ] when followed by a high front vowel [i].

12. /n/ vs /ɲ/

    (a) /men/ (‘bird’) vs /meɲ/ (’nose’)

    (b) /wɐn/ (‘tree’) vs /wɐɲ / (‘cheeky’, ‘scoundrel’)

4.1.1.4 Fricatives

Fricative contrasts occur at two places of articulation: bilabial, and palatal. Both places of articulation show a voicing contrast.

The bilabial fricatives /ɸ/ and /β/ show complementary distribution in some environments. The voiceless bilabial fricative [ɸ] occurs word initially, but not word finally. The voiced [β] does occur word finally, but not word initially. Both occur intervocalically in similar environments.

13. /ɸ/ vs. /β/

    (a) /goɸɑ/ A type of spirit vs. /goβɑ/ An exclamation

    (b) /dæɸi/ 'immigrant' vs. /sæβi/ 'law'

Unlike the bilabial fricatives, the alveolar fricatives /s/ and /z/ have very standard behaviours. There is a voicing contrast, and they can all occur word initially, finally, and intervocalically.

14. /s/ vs. /z/

    (a) /si.te/ 'white; light' vs. /zi.te/ 'afternoon'

    (b) /su/ 'stomach' vs. /zu.zu/ 'rubbish'

The fricative /z/ varies in its realisation, with pronunciations ranging between between [z], [ʒ] and [dʒ]. Some words, such as /zɐβ/ 'hair; fur' is varyingly pronounced as [zɐβ] or [dzɐβ]. 23 out of 32 speakers (76.7%) in the 2017 Nmbo Word List Corpus pronounce /zɐv/ as the fricative [z] with the remaining 9 speakers more [dz]-like in their pronunciations.[3] While unquantified, my impression is that words such as /zi/ 'word; story; language' tend to be realised as [zi] with little community level variation.

Another fricative with variation in its realisations is the glottal fricative /h/. Some Nmbo speakers realise the /h/ [h]-fully, while others drop it almost completely. There are occasional realisations akin to a glottal stop [ʔ]. For example the word /hɑm.bɑ/ 'village' may be realised varyingly as [ʔɑm.bɑ] or [ɑm.bɑ]. This variation is of given /h/ is a Nmbo phoneme that has very clear cognates in its sister languages: e.g. /s/ in Nen, and /ɣ/ in Nama. This variation is the topic of the [h]-drop study (chapter 7).

[3]This was an auditory check combined with a visual inspection of spectograms

4.1.1.5 Trill and Approximants

Nmbo has one trill phoneme /r/, and three approximants: /j, w, l/. /r/ and /l/ are phonemically contrastive, as are /j/ and /w/. All four phonemes can occur word initially, finally, and intervocalically. The approximants can also occur after the vowels /æ, ɐ, o/ to form a diphthong-like sound sequence.

15. /r/ vs. /l/

    (a) /ɑ.reh/ 'to look for' vs. /æ.leh/ 'to hunt'

    (b) /ɸə.rɐh/ 'to become' vs. /ɸə.lɐh/ 'to put something in something'

16. /j/ vs. /w/

    (a) /yɑo/ 'no', NEG vs. /wɑo/ 'ripe'

    (b) /jɐm/ (3sg.nphd/be) /wɐm/ (1sg.nphd/be)

4.1.2 Vowels

Nmbo has eight monophthongal vowels: / i, e, æ,ɑ, o, u / and two central vowels: / ə, ɐ /. The first six vowels I will call *full vowels*, and the latter two *short vowels*. Nmbo also has diphthongs /æy, æw, ɑy, ɑw, eɑ/, and marginal nasal vowels [ẽ] and [æ̃]. The nasal vowels are often used as "non-lexical conversational sounds" (Ward, 2006) such as [ẽː] for affirmation, or "repair initiators" (Dingemanse, Torreira, Enfield 2013) such as [æ̃?]. The exception is the expression *gẽhẽ*, which is used by speakers to draw the attention of the interlocutor to a specific location in the immediate vicinity (section 4.3.3.2).

The full vowels can occur word initially, finally, and function as the nucleus of a syllable. Short vowels have a more limited distribution. Short vowels usually function epenthetically, but can form the nucleus of monosyllabic words (e.g. [wən] 'tree', [k͡pɐβ] 'hole'), and can occur contrastively (e.g. [jəm] 'gumtree' vs [jɐm] the 3sg copula). The behaviour of the short vowels will be discussed in more detail in section 4.1.2.1. The Nmbo vowel space is represented in figure 4.3.

Nmbo does not have phonemic diphthongs. Sequences of phonetic diphthongs end in a high vowel off-glide, and are analysed as a VC sequence (e.g. *yao* [jɑw] 'no', NEG, *kai* [kɑj] 'tomorrow/yesterday'. In the practical orthography, vowel + high vowel sequences are represented as two vowels, e.g. <ai, ao>.

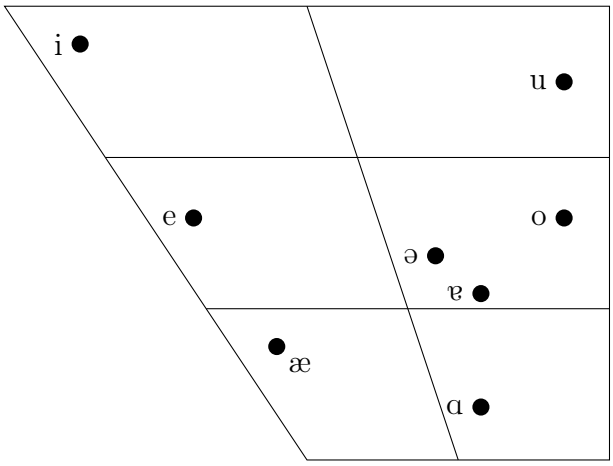

Figure 4.3: Nmbo vowel chart. Placement of vowels based on approximate community mean values of F1 and F2 per vowel, Lobanov transformed. The means are from a subset of speakers in the Nmbo Wordlist Corpus (years 2014 and 2015).

Phonetic diphthongs in a single syllable are very rarely followed by an additional consonant. The words that may have consonants following phonetic diphthongs in the single syllable are either proper nouns (e.g. Kaog [kɑwg], Said [sæjd]), or loan words (e.g. faol [ɸɑwl] ('chicken' from English fowl). The one common noun that looks like an exception is *gaiñ* 'mosquito' and other words that contain it (e.g. *enḡaiñ* a type of fresh water creek fish). These words are analysed as a case of palatal assimilation of the palatal nasal <ñ> [ɲ], giving the monophthong [ɑ] a palatal off-glide. Orthographically, *gaiñ* looks to be a closed syllable with a diphthong nucleus, but it is phonemically /g͡bɑɲ/ and phonetically [gɑʲɲ].

4.1.2.1 Short vowels as phonemic and epenthetic

Nmbo has two central vowels [ə] and [ɐ], which have a shorter duration than the other vowels in its vowel inventory. Similar kinds of vowels have been variously described in other Papuan languages as "transitional vowels" (Foley, 1980), or "predictable vowels" (Blevins & Pawley, 2010). These short, centralised vowels have been attested for all Yam languages which have been described in the last decade. The analysis of their phonemic status, however, differs between the branches. The Nen vowels [ɪ] and [ə] have been identified as mostly epenthetic, but phonemic in some instances with minimal pair examples (Evans and Miller 2016:10). In the Tonda languages these central vowels are analysed as mostly epenthetic Ngkolumpu (Carroll 2016:45-6), but marginally phonemic in Komnzo (Döhler 2018:58), which geographically borders the Nambu branch language of Nama.

I will call the Nmbo [ə] and [ɐ] *short vowels* because a functional term does not capture the full range of these vowels. Terms such as "transitional", "predictable", or "epenthetic", which focus on functions such as syllable repair, do not capture the phonemic element of Nmbo [ə] and [ɐ]. I would argue that the best term of these vowels is not one based on their formal characteristics, but on their physical characteristic; namely that they are significantly shorter in duration than the other six Nmbo vowels (table 4.4). The phonemic and phonetic function of the short vowels may be historically based, as has been analysed for Kalam predictable vowels (Blevins & Pawley, 2010). This is a crucial area to pursue in future research.

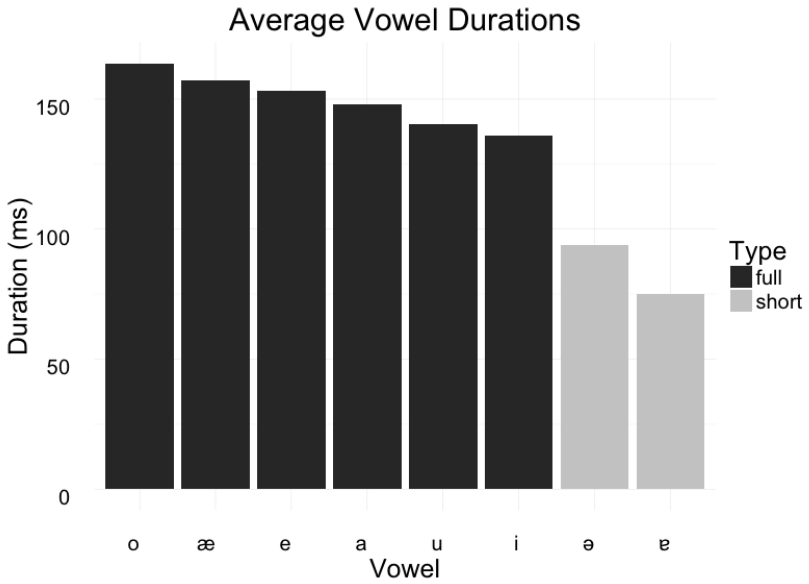

Figure 4.4: Average duration of Nmbo vowels.

Below are some of the characteristics of Nmbo [ə] and [ɐ]:

- They are phonemically contrastive, both with other long vowels, and with each other.
- They can occur word initially in verb roots. They can word finally, but it is unclear whether this is phonemic or merely a tendency for some individuals when releasing final consonants.
- They function as syllable nuclei, and such syllables can take stress.
- Their realisation is not dependent on the quality of the surrounding consonants.
- They break up consonant clusters in syllabification processes.

There are a number of minimal pairs between short and full vowels in Nmbo, and this is the main reason for positing them as distinct phonemes. Minimal pairs between the two short vowels [ə] and [ɐ] have also been attested, warranting two distinct short vowel phonemes.

17. Full vowel vs short vowel

    (a) *sin* /sin/ 'pot' vs. *sén* /sən/ 'tooth'

    (b) *wem* /wem/ Kerake: Yam type; Yarne: 'food', 'yam' vs *wm* /wəm/ 1sg.nphd/be

    (c) *mär* /mæβ/ 'spectator' vs mr /mɐr/ 'brain'

    (d) *muta* /mutɑ/ Yam type vs. *mtar* /mətɑr/ 'quickly'

    (e) *rusa* /rusɑ/ Indonesian loan of 'deer' vs. *rsah* /rəsɑa/ 'to carry'

    (f) *fok* /ɸok/ 'bad omen' vs *fék* /ɸək/ 'loose'

    (g) *hakr* /ha.kr/ 'boy' vs. *hkr* /hɐ.kr/ 'yam cake'

18. /ə/ vs. /ɐ/

    (a) *féd* /ɸəd/ 'edge' vs *fd* /ɸɐd/ 'scar'

    (b) *két* /kət/ grass type vs kt /kət/ 'there'

    (c) *qév* /k͡pəβ/ 'time' vs. *qv* /k͡pɐβ/ 'hole'

19. Full vowel vs. /ə/ vs /ɐ/

    (a) *sov* /soβ/ 'wave' vs *sév* /səβ/ 'hoop' vs *sv* /sɐβ/ 'bundle' vs.

    (b) *yam* /jɑm/ 'custom; way; event' vs. *yém* /jəm/ 'gumtree' vs. *ym* /jɐm/ 3sg.nphd/be

Short vowels can occur word initially as verb roots. When inflected, these short vowel initial verb roots syllabify the vowel with the prefix (table 4.4); the full vowels are completely audible, while the short vowels sound more transitional/epenthetic, or like an insertion. All verb roots that begin with vowels are intransitive verbs in Nmbo, and this also holds for the verb roots that begin with short vowels. Since short vowels occur word initially only as verb roots, it may be that they are performing some kind of phonotactic repair function to maintain the rule that intransitive verb roots begin with vowels.

Infinitive	Gloss	Affixation	Syllabification
avlh	to become crushed	n-avl-n	nav.lɐn
ävrengh	to pack up	n-ävreng-n	näv.ren.gɐn
evoh	to finish	n-evo-n	ne.von
ovarh	to arrive	n-ovar-n	no.va.rɐn
uvätuh	to finish	n-uvätu-n	nu.vä.tun
ɐtorah	to enter	n-tor-yn	nɐ.tor.yɐn
əvermeh	to cross over	n-vrme-n	nə(~ɐ).vɐr.men

Table 4.4: Prefixation and syllabification of verb roots beginning with full vowels and short vowels. The prefixes in these examples are middle prefixes, while the suffixes are all 1SG.A

The function of word final short vowels is less clear. At the end of some monosyllabic words, some speakers produce audible voicing. A similar phenomenon has been observed in Nen, and described as "a voiceless echo" whose quality may differ e.g. not necessarily be [ə] (Evans and Miller 2016:10). So far no minimal pairs have been found.

Short vowels can form the nucleus of a syllable. It has been argued for Komnzo, that the main function of its central vowels is to provide a syllable nucleus where there is none underlyingly (Döhler 2016:64-5). In Nmbo there is little doubt that phonetically the short vowels are present, but I will also argue that they are not there underlyingly. The analysis is the same as Döhler's, and I present syllabification rules as evidence. For example when the verb root *gmeh* [gɐ.meh] 'to hit' is inflected with the 3sg undergoer prefix of the ɑ-set, the form becomes *ygmetan* [jɐg.me.tan] 'I hit it'. The *-tan* is the 1sg actor suffix. The onset [g] of the verb root syllabifies with the *y-* prefix, and becomes the first syllable [jɐg]. The [ɐ] of the first syllable in the root disappears, which suggests that [ɐ] is not underlyingly present. The short vowel [ɐ] is functioning purely as a syllable nucleus to maintain a licit syllable structure. Examples of the syllabification process are presented a bit further on (section 4.1.3.2).

Despite this epenthetic vowel-like behaviour, the short vowels exhibit other behaviours that are atypical of epenthetic vowels. Firstly, syllables with a short vowel as a nucleus appear to take stress. Nmbo stress is obligatory, and tends to fall on the penultimate syllable. Many multisyllabic Nmbo words that take stress do not contain full vowels, e.g. *dqn* [ˈdə.k͡pɐn] 'black anthill', *kkv* [ˈkɐ.kɐβ]

'garden', *mlml* ['mɐl.mɐl] 'sweat'. Secondly, the realisations of the short vowels do not appear dependent on the quality of the surrounding consonants. Figure 4.5 shows a slice of natural speech *mna kt* 'because there', and the voicing is visible between the two voiceless stops [k^h] and [t^h]. Note also the short vowel [ə] between [m] and [n] in the word *mna*.

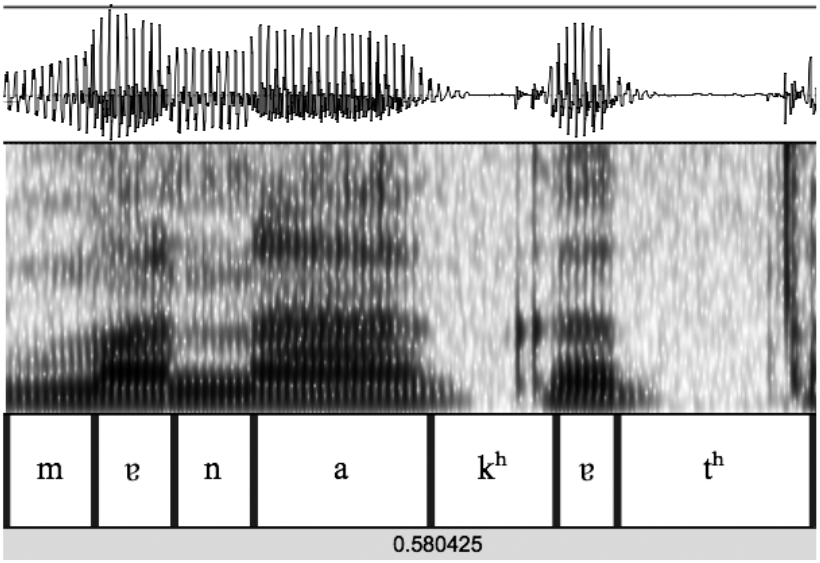

Figure 4.5: Spectogram of *kt* 'there'. WSEK1-B20150817-02DimbanKSae, 00:05:54.

The phonological status of [ə] and [ɐ] thus appears to be two-fold: to function as a fully-fledged phonemic vowel in some instances, and to function as a phonologically visible epenthetic vowel in others. Evans and Miller (2016) analyse the Nen [ɪ] and [ə] as phonemic and epenthetic, however most of the time both central vowels are epenthetic (2016:10). The statement is also true for Nmbo.

	Word/Syllable Initial	Nucleus of closed syllable	Word /Syllable Final
i	*ide-ide* (a kind of tree)	*wim* ('scent')	*zi* ('language')
e	*eg* ('fog')	*men* ('bird')	*de* (honeyeaters)
æ	*ær* ('person', 'man')	*tær* (friend)	*bæ* (3ABS)
u	*ur* ('bush fowl')	*gu^{m}b* ('swamp')	*gu* (Great Billed Heron)
o	*o^{n}da* ('dream')	*non* ('why')	*daro* ('butterfly')
ɑ	*ag* ('coconut')	*kal* ('wound')	*ka* ('where')
ə	*əvermeh* 'to cross over'	*wən* ('tree')	?
ɐ	*áwih* (to fall)	*yɐm* (3sg.nphd/be)	?

Table 4.5: Occurrences of Nmbo vowels in word initial, medial, and final position.

4.1.3 Phonotactics

4.1.3.1 Syllable Structure

Maximal syllable structure in Nmbo is $[CCVC]_\sigma$ (diagram 4.1), and the minimal syllable structure is $[V]_\sigma$. Both structures have restrictions and specifications in their occurrences, which are outlined below:

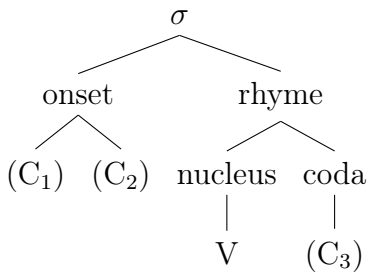

Diagram 4.1: Structure of Nmbo syllable.

- An onset consonant by itself (C_1) can be any consonant available in the inventory, except when word initial. These follow the rules of consonant occurrences (see section phon:consonants):
  - /β/ cannot occur word initially, but can form the onset of word medial syllables.
  - Prenasalised stops cannot occur word initially, but can form the onset of word medial syllables.
- There are restrictions as to which consonants can form a consonant cluster (C_1C_2). Namely, C_1 must be a voiced, voiceless, or prenasalised stop, or any fricative except /h/. C_2 must be /r/. See table A.1 for licit tautosyllabic clusters.
  - Licit CC clusters occur quite freely as onsets of word initial syllables, but seem less frequent as onsets of word medial syllables. There may be further restrictions, but these have not been investigated.
- The coda of a syllable must be simple (C_3) e.g., **.tekr* but *te.kr*. The possible consonants in coda position are limited due to the restriction of heterosyllabic consonant clusters.[4]

[4]The possibility of a rhotic + prenasalised stop was considered as a possible complex consonant cluster allowed in coda, e.g. <trmb> 'flower for fruiting plants', <trnd> 'sandpaper tree'. There is, however, no phonetic, phonological, or morphophonemic evidence strongly in favour of or against this possibility. Rather than include an additional complexity on weak

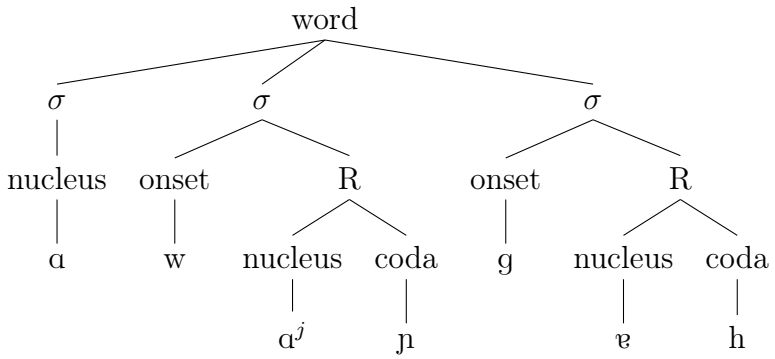

Figure 4.6: Syllable structure for orthographically diphthongal <awaingh>

- V may be any vowel in the vowel inventory. This includes short vowels that are phonemically present as well as epenthetically inserted.

The minimal syllable structure of $[V]_\sigma$ can occur in any position of a word, but this distribution is restricted. $[V]_\sigma$ can occur word initially for uninflected lexical words, but does not seem to occur medially or finally. $[V]_\sigma$ can, however, occur word medially and finally in inflected words.

For a word medial example, *namseaend* [nɑm.se.**ɑ**.eⁿd] is the verb root *ämsuh* [æmsuh] ('to sit') inflected for a 3du sole argument in the perfective preterite tense. The desinence *-eaend* does not break the vowel sequences with a palatal glide, and results in the [ɑ] as a single syllable. The same occurs with other verb roots inflected for dual number and perfective preterite tense, such as *no.vn.ge.**a**.end* ('those two began'), *y.nn.dan.ge.**a**.em* ('we two opened it'), *ym.de.**a**.em* ('we two told him').

Word final example of $[V]_\sigma$ is also from an inflected word. The dative *-e* suffixed on certain words can be phonetically realised as [e] without a palatal glide preceding it, e.g. *foa-e* [fo.wa.**e**] (later-DAT). Verb roots with final open syllables can also take the dative suffix, resulting in forms such as *roha-e* [ro.hɑ.**e**] 'draggingly'.

evidence, I have chosen to keep the structure of the syllable simple and analyse /r+prenasal/ clusters are being broken apart via epenthesis into two syllables.

	Word initial	Word Medial	Word Final
V	u.rer ‘trap type’	nɑm.se.**ɑ**.e^{n}d*	ɸo.wɑ.**e**
	ɑ.hoh ‘to get dressed’		
	eno ‘female personal name’		
VC	eg (‘fog’)	NA	NA
	or ‘easterly wind’		
CV	hi ‘bush torch’		k͡pɐ.ki (‘container’)
	kɑ‘where’		de.βe ‘father’
	gu ‘Great-billed heron’	jən.ɸi.jɑk (Whistling Kite)	sɑw ɑ‘cane’
CVC	tok ‘top; above’	o.tam.beh ‘to race’	son.zær ‘pebble’
	dæβ‘when’	ku.ræm.gu place name	hɑ.rɐk͡p‘yam sapling’
	kɑj ‘tomorrow’		gæ. ræz ‘pandanus’
	sɑw.ɑ‘cane’		dæn.dæj ‘always’

Table 4.6: Occurrence of possible syllable types per word position.
‘*’ = “those two sat”

4.1.3.2 Syllabification

The main process of syllabification is the mapping of strings of phones onto licit syllable structures. The order in which syllabification maps a syllable template is akin to that of Komnzo, a distant sister language of Nmbo’s in the Tonda branch of the Yam languages (Döhler 2018:68-71). The procedure for syllabification goes from right to left as follows:

1. Associate each specific vowel with a syllable nucleus.
2. Identify any prenasalised stops as an overlap phone. Prenasalised phones will always be assigned as an onset and coda, unless they are word final. Assign accordingly (which also means adding a short vowel as nucleus after the prenasalised stop).
3. Establish and maximise onset in accordance with syllable template.
4. From right to left: Break up unsyllabified consonants with short vowels
5. Assign rhyme.

Tables 4.7 and 4.8 demonstrate how the rule is implemented with concrete examples.

/frkta /	Underlying representation
frkt[a]	Step 1: Associate each specific vowel with a syllable nucleus.
frk[ta]	Step 3: Establish and maximise onset in accordance with syllable template.
[frɐk] [ta]	Step 4: From right to left: Break up unsyllabified consonants with short vowels.

Table 4.7: Syllabification process on word, example 1

/qakiyongrwn /	Underlying representation
q [a] k[i] y [o] ngrwn	Step 1: Associate each specific vowel with a syllable nucleus.
q [a] k[i] y [o] [n.gɐ] rwn	Step 2: Identify any prenasalised stops as an overlap phone.
[qa] [ki] [yon.gɐ] rwn	Step 3: Establish and maximise onset in accordance with syllable template.
[qa] [ki] [yon.gɐ] r [wɐn]	Step 4: From right to left: Break up unsyllabified consonants with short vowels.
[qa] [ki] [yon.gɐr] [wɐn]	Step 5: Assign coda.

Table 4.8: Syllabification process on word, example 2

4.1.4 Intonation

Pitch, intensity, and duration are the primary manifestations of intonation that have been identified cross-linguistically. Pitch is clearly significant for distinguishing between intonation types in Nmbo. In this section I will describe intonation types for declaratives and interrogatives, and the characteristic of their boundary markers. Intonation types are analysed as applied over intonation units (IUs), and intonation units are demarcated by a type of IU boundary marker.

Declarative utterances show individual speaker variability in their pitch ranges. Figure 4.7 is an instance of a relatively vivid pitch range applied across a declarative utterance. This declarative intonation type shows a relatively consistent rise and fall through most of the utterance, showing declination until it clearly drops towards the end in an instance of final lowering. The pitch drop clearly marks an utterance *final fall boundary marker*. This boundary marker signals the end of a speaker's turn, and often marks the end of a discourse topic.

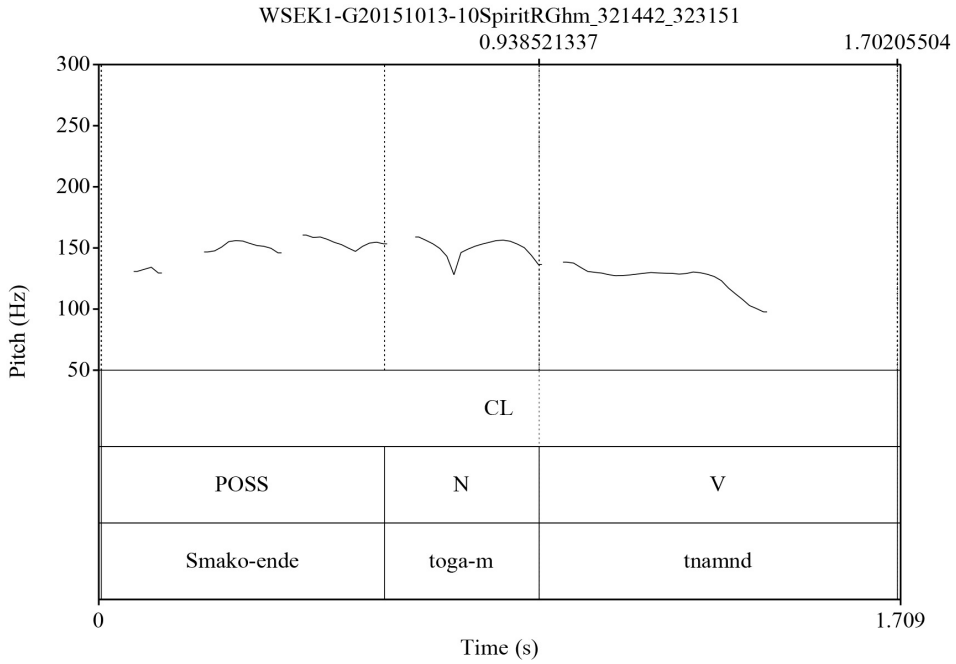

Figure 4.7: Pitch contour of intonation unit, marked by a final fall boundary marker. The utterance can be translated as "Smako's child shot [the animal]."
(33 y.o male, G20151013-10SpiritRGhm 321442 - 323151)

Contrast this utterance with figure 4.8 showing a flat pitch with little range, but still sits at the end of a discourse topic. This is an instance of a *final flat boundary marker*. The clause following this particular utterance changes the

discourse topic, which suggests the flat boundary marker can function similarly to a falling marker. The key observation is that changes in discourse topic are not always identifiable by a fall in pitch in Nmbo. Flat markers are also more common in long strings of utterances that have little to no pause between IUs.

Figure 4.9 is a pitch contour of a *continuing boundary* marked utterance from the same speaker in the same recording as the final flat example. This utterance cannot be taken as the end of a conversational topic, and the boundary cannot be interpreted as demonstrating a final pitch drop. This utterance (example 6a) is followed by a pause of about a second long before continuing with the topic of hot water (6b) :

(6) a. nu bés, y-n\eho/sm... (1sec)

water fire 1nsg.A>3sgU:pfv.pst:ven/put

We put the hot water on the fire...

b. nu bés-e t\ne/ham Edi....

water fire-DAT 1nsg.A>3sgU:ipfv.ypst/wash Eddy

We washed Eddy with hot water.

(WSEK1-A20160722-04RM01Ypsthm, 00:03:09.902 - 00:03:14.131)

Perceptually, and by crude visual assessment of these pitch contours, the final flat and continuing markers seem very similar in terms of relative pitch movement against the rest of the intonation contour. I do not know whether these two boundary markers are perceptually different for Nmbo speakers, but I have found it challenging to distinguish the two types. For the purposes of this sketch grammar I have described these boundary markers as functionally distinct, but it may be that the acoustic and prosodic characteristic of these boundary markers are, in fact, of the same kind.

As a final observation on falling boundary markers, it is worth noting that the final fall appears much more common when the clause final element is a subordinate clause, or an adjunct. Additionally, it is common for a major downstep in pitch between main clauses and adjuncts, with the pitch contour over the adjunct being significantly lower than that of the main clause (figure 4.10).

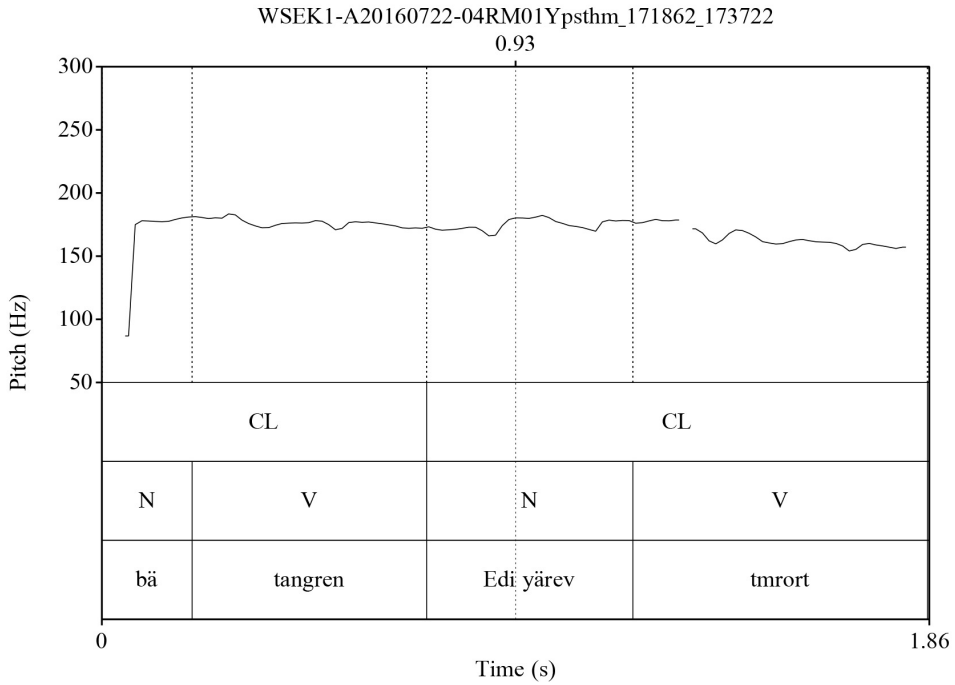

Figure 4.8: Pitch contour of intonation unit, marked by a flat final marker at the end. Notice the relative lack of pitch range across the entire IU. The utterance can be translated as "Those two argued with Eddy's father." (30 y.o female, A20160722-04RM01Ypsthm , 00:02:51.136 - 00:02:53.673)

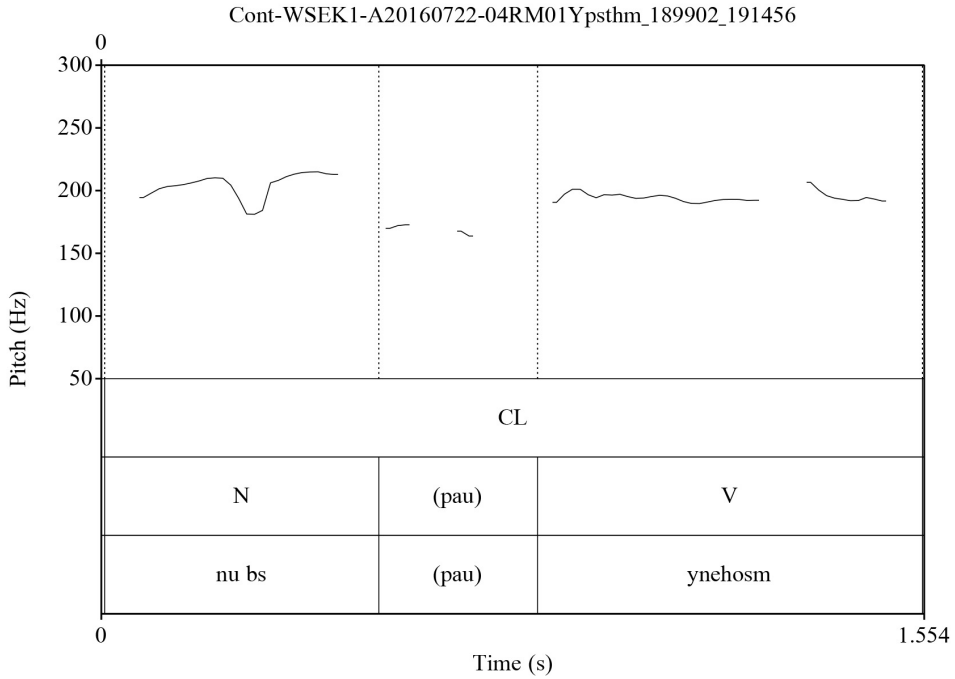

Figure 4.9: Pitch contour of intonation unit, marked by a continuing boundary marker ar the end. The utterance can be translated as "The hot water, we boiled it." (30 y.o female, A20160722-04RM01Ypsthm, 00:03:09.902 - 00:03:14.131)

Nmbo speakers often produce a *final raising boundary marker*. Figure 4.11 shows the pitch of a raised final boundary at the end of the clause. In the original narrative, there is a pause of approximately four seconds before the speaker continues on with the narrative, but the new discourse topic is different to that of this particular utterance. Impressionistically, it seems that raised are used at the end of utterances to signify some sense of temporal passing.

Questioning intonations in Nmbo often do not have a rising boundary. Both polar and content questions have a downward trajectory of F0 across the clause with a final falling boundary marker at the end of the verb (figure 4.12 for a polar question, figure 4.13 for content question). This lack of a falling question boundary marker has been observed in other parts of Southern New Guinea (e.g. Gregor pers. comm. for Yelmek).

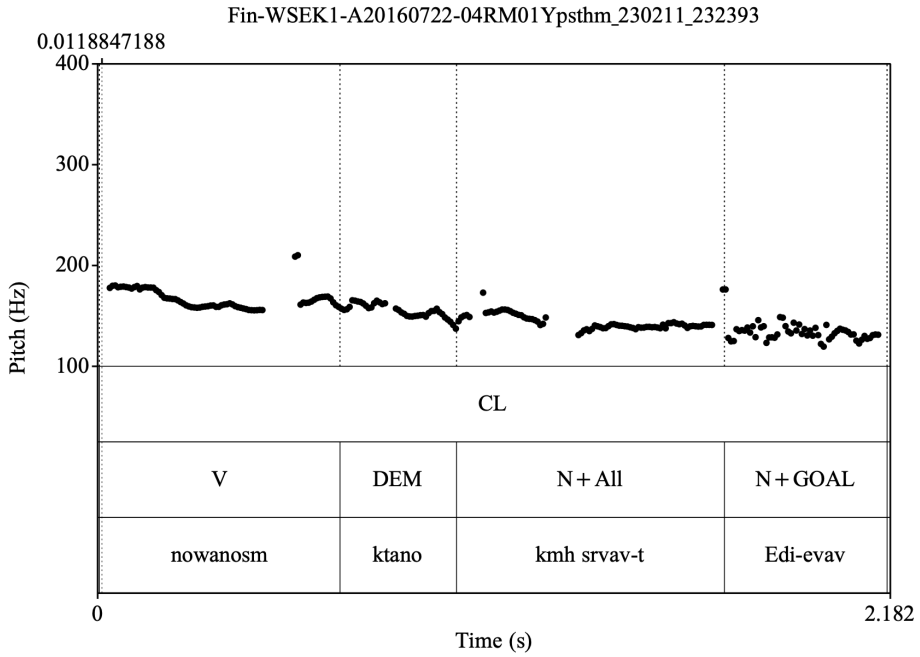

Figure 4.10: Pitch contour of intonation unit, marked by a raising final boundary marker. Utterance by a thirty year old woman from Bevdvn. Translation of utterance: "We arrived there to the sleeping place, to Eddy." (WSEK1-A20160722-04RM01Ypsthm)

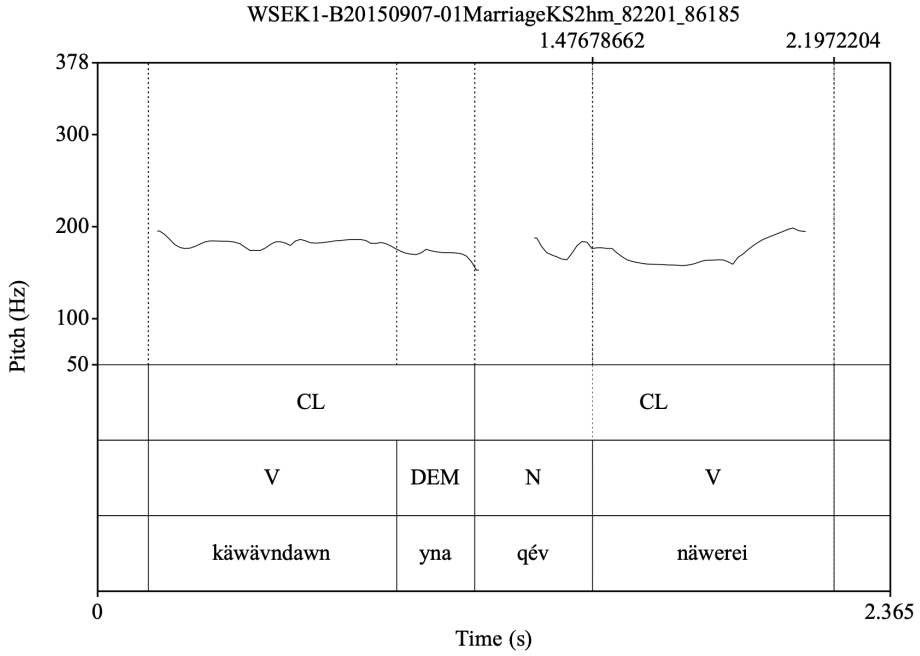

Figure 4.11: Pitch contour of intonation unit, marked by a raised final marker. The utterance can be translated as "I was waiting, in darkness she came". 35 y.o male, B20150907-01MarriageKS2hm, 00:01:24 - 00:01:26)

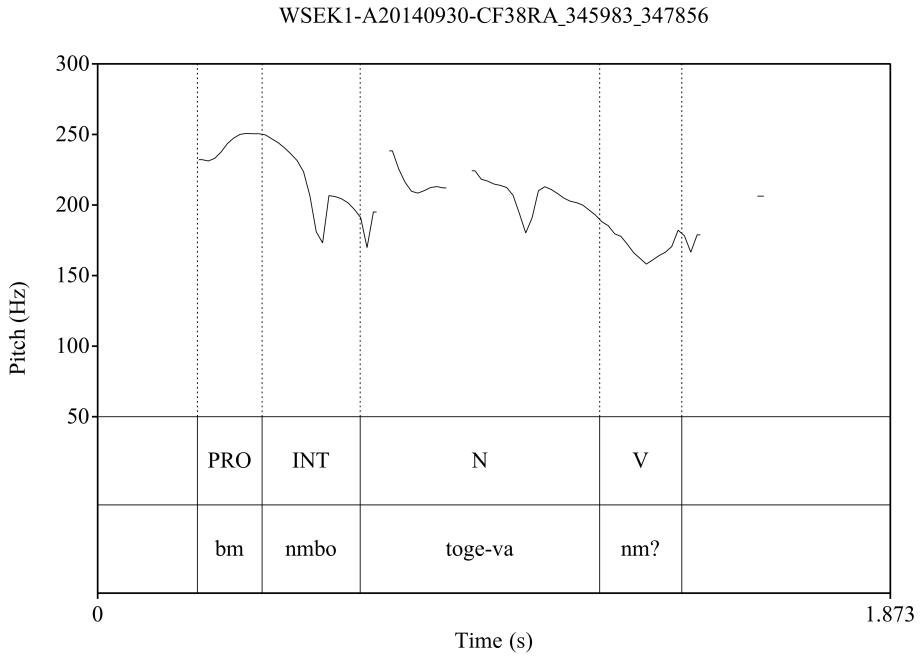

Figure 4.12: Pitch contour of a polar question clause, "Do you have children?" (30 y.o female, WSEK1-G20170617-01Alqi02MQhm)

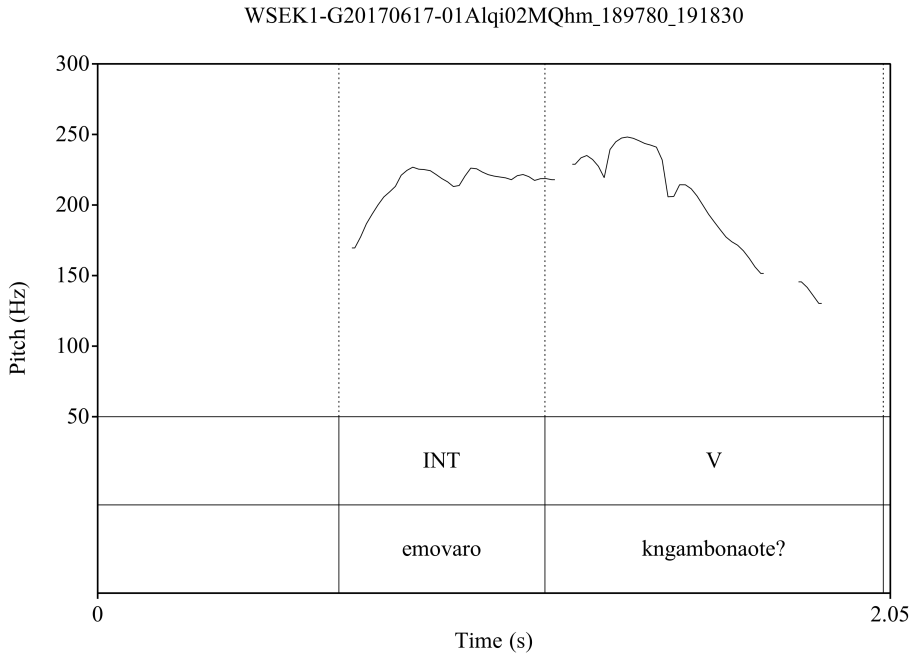

Figure 4.13: Pitch contour of a content question clause, "Who did you play with [when you were a child]?". INT = interrogative pronoun (30 y.o female, WSEK1-G20170617-01Alqi02MQhm).

4.2 Word Classes

Nmbo has three major word classes: nominals (section 4.2.1), verbs (section 4.2.2), and particles (section 4.2.3). Each class has multiple subsets with their particular distributional and morphosyntactic properties. The primary distinction between the three major classes are that nominals and verbs inflect, while particles do not. Discourse function markers and typical collocations will briefly be listed in section 4.2.4.

4.2.1 Nominals

Nominals are a large and diverse class of words whose condition of inclusion in this category is defined thus:

- Morphologically: They part-take in a paradigm of grammatical cases, and/or locational cases, and/or other semantic cases. For a discussion on different case types see section 4.4.4).
- Syntactically: They can form a head of a noun phrase, and/or they can modify the head of a noun phrase.
- Semantically: They prototypically denote humans, non-human entities and objects, locations, and temporal concepts. Abstract concepts such as 'desire', or descriptors such as 'large', 'red' are also included.

There are seven subclasses of nominals: personal pronouns (section 4.2.1.1), common nouns (section 4.2.1.2), demonstratives (section 4.2.1.3), adjectives (section 4.2.1.4), quantifiers and numerals (section 4.2.1.5), interrogatives (section 4.2.1.6), and indefinite pronouns (section 4.2.1.7). Common nouns are open class, while the remaining types are closed class. Common nouns can be further distinguished into the semantic types of human, non-human, location, and temporal. Infinitive verbs behave like common nouns, and their behaviours and characteristics will be described in the infinitive marker section (section 4.4.1) of the nominal morphology section.

4.2.1.1 Personal Pronouns

Personal pronouns are a phonologically free, closed class of words that distinguish three persons in both singular and non-singular number. They cannot be modified by demonstratives or adjectives. The function of personal pronouns

is primarily dictated by their morphology; that is, whether they are expressing grammatical case, are in the possessive form, are taking locational morphology, or expressing other semantic roles.

The pronominal expression of core grammatical relations is carried out by the ergative, absolutive, and dative forms (table 4.9). All other personal pronouns expressing semantic relations are built up from either the absolutive or dative forms. Since the semantics expressed by the pronouns of semantic relations are transparent and compositional from their morphology, they will not be covered here (see section 4.4.4 for semantic morphemes and their functions). An overview of all known pronominal forms are shown in table 4.11.

Person	Absolutive	Person Number	Ergative	Dative
1	ynd	1sg	yndo(n)	ta
		1nsg	yndvem	tve
2	bm	2sg	bmo(n)	be
		2nsg	bmovem	bve
3	ba	3sg	ymo(n)	ya
		3nsg	ymovem	yve

Table 4.9: Core grammatical pronouns: absolutive, ergative, and dative forms.

This section will describe the basic absolutive, ergative, and dative forms of the pronoun, as well as the close possessive and 'alone' forms which are unique to personal pronouns. For interrogative pronouns, see section 4.2.1.6.

The absolutive pronoun represents the sole argument of a monovalent predicate. First, second, and third person categories are expressed, while number is not expressed at all. In the examples below, the number of the absolutive referent is shown on the relevant morpheme of the verb.

(7) a. [Singular argument.]

ynd **w-**n/**m**

1ABS **1sg.U**:nphd/come

I'm coming.

b. [Dual argument.]

ynd **yn-**n/**ren**

1ABS **1du.U**:nphd/come

We two are coming.

c. [Plural argument.]

ynd **yn-**n/**m**

1ABS **1pl.U**:nphd/come

We (plural) are coming.

The ergative pronoun denotes the agent of di- or trivalent predicates. Unlike the absolutive forms, the ergative makes a singular vs non-singular number distinction across three person categories. The label *non-singular* is used to capture how dual and plural arguments are expressed by the same forms, while the singular form stands distinct. Dual number is therefore not distinguished on the independent pronouns, but is encoded on the verb complex.

The singular forms of the ergative pronouns have a variant with a final [n] (shown in table 4.9). Speakers do not report any semantic difference between the two forms.

(8) a. [Singular A]

yndo bä y\mi/tan

1sg.ERG 3ABS 1sg.A>3sg.U:ipfv.nphd/ask

I am asking him.

b. [Dual A]

yndvem bä y\mih/**wm**

1nsg.ERG 3ABS **1du.A**>3sg.U:ipfv.nphd/ask

We two are asking him.

c. [Plural A]

yndvem	bä	y\mih/**tam**
1nsg.ERG	3ABS	**1pl.A**>3sg.U:ipfv.nphd/ask

We (plural) are asking him.

The dative pronouns also distinguish singular and non-singular number for the three person categories. The dative is one of the three pronominal forms used for expressing the third argument of a three-place predicate (the other two being the benefactive/purposive, and goal forms of the pronoun). Example 9a shows the recipient in the dative form of the pronoun, while (b) shows the goal pronoun which is built up from the dative form with the goal clitic *=vav*. The verbal indexing of the recipient argument shows variation, and will be covered in the trivalent clauses section 4.8.3.

(9) a.

Ymo	ämbru	ynḡ	w-a-ramo-i	**ta**.
3sg.ERG	one	bag.ABS	1sg.U-APP-give-1sg.A:pfv.pret	**1sg.DAT**

He gave me one bag.

b.

Ymo	ämbru	ynḡ	y-ramo-i	**ta-vav**.
3sg.ERG	one	bag.ABS	3sg.U-give-3sg.A:pfv.pret	**1sg.DAT=GOAL**

He gave me one bag.

Fieldnotes 2017 Book2, p.9

Personal pronouns in the dative form, and sometimes the absolutive form, can take a variety of semantic case markers (section 4.4.4). We will discuss most of these built-up pronouns in the nominal morphology section, since the function of the semantic case markers are similar to when building on common nouns. For the remainder of this section we will discuss two forms that are built by morphemes that are unique to personal pronouns: the *close possessive -nzo* and the *alone -^{m}bia* suffixes.

The *close possessive* is a term used by Döhler (2018:145-5) in reference to a prefix found in Komnzo.[5] It refers to a non-obligatory marking of possession

[5]Evans describes the Nen equivalent as the *reflexive possessive* since these pronouns can be used in reflexive/reciprocal constructions as the core argument. This has not been tested for

that is often used to mark kin terms, e.g. *ta-nzo ama*, 'my *real* mother'. Nmbo speakers explain the distinction in use between the regular and close possessive as akin to saying "really mine". The close possessive pronoun can be used to indicate possession of any nominal; it is not limited to kin terms. The use of the close possessive pronoun is not obligatory, and some classically inalienable things such as body parts are rarely if ever possessed using it. Example 10 shows a contrastive example of the close possessive pronoun (a), and the regular possessive pronoun (b).

(10) a. [Alqi is talking about language use in her household when she was a child.]

Tanzo	ama-m	**yä-nzu**	zi-e	q-ng\mde/tao
1sg.POSSC	mother-ERG	3sg.POSSC	language-DAT	3sg.A>1sg.U:ipfv.rmpst/talk

My own mother talked to me in her own language.

WSEK1-G20170617-01Alqi02MQ, 00:07:01.797 - 00:07:06.763

b.

Tande	ama-m	ynd	Arovwe=n	w\rav/tai
1sg.POSS	mother-ERG	1ABS	Arovwe=LOC	3sg.A>3sg.U:pfv.pret/give.birth

My mother gave birth to me at Arovwe.

WSEK1-A20140930-CF38RA, 00:03:20.840 - 00:03:22.920

The *alone suffix* -^{m}*bia* expresses a state whereby a person is described as being "all alone". The dative form of the pronoun forms the root. Neither the alone suffix, nor any other bound morpheme resembling it in form, have been observed on any other nominal root. Note in examples 11 that the alone pronoun is used attributively, but agrees with the person and number of the sole argument. The alone pronoun itself does not have the status of an independent argument.

Nmbo, and as I have no evidence to support possible reciprocal uses for these pronouns I have chosen the semantic and functionally based label used by Döhler.

(11) a. Ynd **ta-mbia** q-ng/maro

1ABS **1sg.DAT-ALN** 1sg.A:rmpst/go

I alone went by myself.

WSEK1-A20170613-03Rhouda06Intro, 00:00:53.650 - 00:00:54.827

b. Ynd **ta-mbia** zi k\oav/tawn kt!

1ABS **1sg.DAT-ALN** word.ABS 1sg.A:ipfv.rmpst/speak DEM.dist

I was all by myself, talking there!

WSEK1-B20150813-06DingyHZ, 00:03:18.771 - 00:03:20.963

4.2.1.2 Common Nouns

Common nouns are an open class of nominals that can function as heads of noun phrases. There are four subsets based on slight differences in morphological behaviour: human, general, locational, and temporal nouns (table 4.10).

Subtype	Example
Human Nouns	*är* 'man', *äusa* 'old woman', *toge* 'child'
Non-human Nouns	*aha* 'dog', *nne* 'food', *yäme* 'mat'
Locational Nouns	*hamba* 'village', *mngo* 'house', *orng* 'jungle'
Temporal Nouns	*foa* 'later', *kai* 'tomorrow/yesterday', *qvte* 'night'

Table 4.10: Subtype of common nouns, with examples

The subsets are identified by the types of case markers they take. For example, human common nouns do not take the allative *=t* when in the semantic role of GOAL, but will take a dative *-e* followed by the goal *=vav*. Likewise locational nouns can host the allative when they are in the semantic role of GOAL, but they cannot take the goal morpheme *=vav*.

Human and non-human common nouns behave similarly. They function as heads of noun phrases, and often take the semantic macro-role of *actor* or *undergoer*. They both take grammatical case markers like the ergative *-m*. Non-human nouns cannot, however, take the plural form of the ergative case marker *-vem*, while human common nouns can (examples 12 and 13). The only other case that makes a non-singular number distinction is the dative (section 4.4.4).

	Host	Clitic	1sg	1nsg	2sg	2nsg	3sg	3nsg
Ergative			yndo(n)	yndvem	bmo	bmovem	ymo	ymovem
Absolutive	ABS	(zero)	ynd	ynd	bm	bm	bä	bä
Dative	DAT		ta	tve	be	bve	yä	yve
Possessive	DAT	-nde	ta-nde	tve-nde	be-nde	be-nde	yä-nde	yve-nde
Close Poss.	DAT	-nzo	ta-nzo	tve-nzo	be-nzo	bve-nzo	yä-nzo	yve-nzo
Goal	DAT		ta=vav	tve=vav	be=vav	bve=vav	yä=vav	yve=vav
Benefactive	DAT		ta=wt	tve=wt	be=wt	bve=wt	yä=wt	yve=wt
Locative	DAT	-vav-n	ta-vav-n	tve-vav-n	be-vav-n	bve-vav-n	yä-vav-n	yve-väv-n
Allative	DAT	-vav=t	ta-vav=t	tve-vav=t	be-vav=t	bve-vav=t	yä-vav=t	yve-väv=t
Ablative	DAT	-vav=we	ta-vav=we	tve-vav=we	be-vav=we	bve-vav=we	yä=vav=we	yve=vav=we
Comitative	ABS/ERG	=va	ynd=va	yndvem=va	bmo=va	bmovem=va	ymo=va	ymovem=va
Privative*	ABS	=ovnär	ynd=ovnär*	-	-	-	-	-
Originative	DAT		ta-vav=mn	tve-vav=mn	be-vav=mn	bve-vav=mn	yä-vav=mn	yve-väv=mn
Additional	Host	Morpheme						
Restrictive	ABS	=ro	ynd=ro	UA	UA	UA	UA	UA
Only	ABS, ERG	=yo	ynd=yo	UA	UA	UA	UA	UA
Alone	DAT	-mbia	ta-mbia	tve-mbia	be-mbia	bve-mbia	yä-mbia	yve-mbia

Table 4.11: Elicited pronominal paradigm. * = Offered by Govav speakers, rejected by Bevdvn/Arovwe speakers. UA = Unattested.

(12) a. yna rokar gym, är**-m** y\avro/tat.

DEM.prox thing [TOPIC] man-ERG 3sg.A>3sg.U:ipfv.nphd/do

This object [TOPIC] a man made it.

WSEK1-G20170624-03Dmbom1, 00:01:57 - 00:02:01

b. sombwi är**-vem** nndkv=an q\vyo/nge...

two man**-ERG.nsg** middle=LOC 3duA>1sg.U:pfv.pret/carry

Two men carried me in the middle.

WSEK1-G20150902-03PigGS, 00:11:52.780 - 00:11:56.340

(13) a. [Two dogs are chasing a pig. The ergative *-m* is the basic singular form used on human referents, while the verb indicates a dual agent argument.]

mamwi k-ng\rsa/o, aha**-m** den mé

pig.ABS 3sg.A:ipfv.rmpst:ven/run dog**-ERG** ALR CONT

t-ng\anam/**wt**...

3du.A>3sg.U:ipfv.rmpst:and/chase

The pig ran away, the two dogs continued to chase it...

WSEK1-B20141007-PigYS,00:17:30.236 - 00:17:31.546

Some human common nouns can also express plural number by reduplication (e.g. *mrz* 'girl', *mrz-mrz* 'girls' ; *hakr* 'boy', *hakr-hakr* 'boys'). Reduplication of non-human common nouns does occur, but often fulfils different functions depending on the root being reduplicated, such as forming a plant name from an animal name or vice-versa (section 4.4.3).

Locational nouns are common nouns whose semantics inherently imply a location; e.g. *hamba* 'village', *mngo* 'house', *orng* 'jungle'. Locational nouns take locational semantic case markers that human nouns do not; these are the *locative*, *allative*, and *perlative* suffixes (sections 4.4.4.6 - 4.4.4.9). In other words locational nouns do not function as core grammatical arguments S, A, or P, but are heads of complement/adjunct NPs.

(14) a. Gnoso kt y/m **mngo**=n.

now DEM.dist 3sg.be/nphd **house**=LOC

Now she is there, at home.

WSEK1-G20150921-LostZS, 00:03:52.335 - 00:03:53.933

b. Bä **mngo**=t ä n-ng\ang/ot.

3ABS **house**=ALL FUT 3sg.A:ipfv.nphd:and/return.back

She will go back to the house.

WSEK1-B20150808-AuyaYrnAT, 00:04:37.709 - 00:04:39.259

Temporals are a functional class of nouns with temporal semantics (Döhler 2018:97 for Komnzo). They are the only class of words that can take the temporal morphemes *=tawa\=tao* (example 15). Temporal nouns otherwise behave much like locational nouns, in that they can take locational case markers to indicate that a particular event takes place at that time (e.g. the locative on the word 'morning', *kai yägu=n* 'at tomorrow/yesterday morning'). Table 4.12 shows a list of common temporal nouns.

Form	Meaning
evh	'day', 'time'
dändäi	'always', 'a long time ago'
däv	'at that time', 'when'
gnoso	'now'
kai	± 1 day (i.e. 'yesterday' or 'tomorrow')
nambt	±2 days (i.e. 'two days ago' or 'two days later')
yägu	morning
zite	afternoon
*taim**	time

Table 4.12: Temporal nouns. Note the English loan word *time* is now commonly used as a temporal noun, often in the expression *ämb taiman* 'at some time' (example 18)

(15) a. nne mna t\aneh/an, zite=**tawa**.
food.ABS DEM.fw 1sg.A>3pl.U:ipfv.ypst/cook afternoon=**TEMP**
I was cooking food, during the afternoon.
WSEK1-B20170720-02Yayam01YPst, 00:01:00.116 - 00:01:02.37

Temporals can modify head nouns by functioning attributively. When they do, they form a constituent with the noun phrase.

(16) [Ynd bä q-ng/m] [**kai** yägu=n.]
1ABS FUT 1sg.U:nphd/go ± **1day** morning=LOC
I will go tomorrow morning.

(17) a. tande endne mñutyu n\evo/yng. **Däv**
1sg.POSS sickness.ABS all 3sg.A:pfv.pret/finish **when**
Daru=t ge n\ovar/yn.
Daru=ALL DEM 1sg.A:pfv.pret/arrive
My sickness completely ended. When I arrived to Daru.
WSEK1-B20150818-01DaruTT, 00:02:00.125 - 00:02:04.563

b. [Alqi is explaining when and why she speaks languages other than Nmbo. The following utterance is in relation to her mother's language, Neme:]

Ama-nde zi **däv** n\owav/tan dena,
mother-POSS language.ABS **when** 1sg.A:ipgv.nphd/speak DEM.mn
ama-nde hakr ynano ge y-n/m.
mother-POSS brother.ABS here.LOC DEM 3sg.U:nphd/come
I speak my mother's language when, for example, mother's brothers come here.
WSEK1-G20170617-01Alqi02MQ, 00:08:24.803 - 00:08:28.745

(18) [An example of the loan word *taim* 'time' being used as a temporal noun. Notice the speaker immediately corrects the expression to use the indigenous Nmbo word *evh* 'day' in the same construction.]

ämb	**taim=an**	yndvem,	ah,	**ämb**	**evh=on**	ta\mndo/ngm
Some	**time**=LOC	1nsg.ERG	DISC	**some**	**day=LOC**	1pl.A>3pl.U:pfv.fut/tell

dena...

QUOT

Some times, ah, some days we tell [our wives] like the following...

WSEK1-B20150908-04PerceptualKS, 00:12:06.200-00:12:09.900

4.2.1.3 Demonstratives

Demonstratives are deictic expressions that serve specific syntactic functions (Diessel 1999:2). Nmbo has two kinds of demonstratives. One set are the nominal demonstratives that take semantic case marking. These demonstratives do not take grammatical case markers, and cannot be possessed. These different subsets of demonstratives can take other kinds of case markers.

Nmbo also has a set of adverbial demonstrative particles *ge* and *gs*, which does not take any inflectional morphology. These are presented in section 4.3.3.2 under the particles section.

Form	Label	Functions
yna	proximal	pronominal, locational, adnominal
kt	distal	pronominal, locational, adnominal
dena	manner	manner
mna	manner:forward	anticipatory/place holding

Table 4.13: Nmbo demonstratives.

Nmbo has two subsets of demonstratives with multiple functions (summarised table 4.13). The pronominal *yna* and *kt* make a proximate and distal distinction and are glossed as such. I have tentatively glossed *dena* as a general 'manner demonstrative', while *mna* is glossed as a 'forward looking demonstrative'.

Yna and *kt* have multiple functions. Firstly, the root forms can function pronominally. They function anaphorically and stand in for entities and/or discourse blocks which were mentioned previously:

(19) a. **yna** ädi n\evo/.
DEM.prox INTS 3sg.A:pfv.pst/finish
This [the story I just told] is truly finished.
WSEK1-G20170724-01FY11Fhm, 00:05:10.014 - 00:05:12.219

b. foyo, **kt** ädi n\evo/.
then **DEM.dist** INTS 3sg.A:pfv.pst/finish
So, now that's truly finished.
WSEK1-KaratPig, 00:02:51.738 - 00:02:52.987

Yna and *kt* also function as locational demonstratives. *Yna* is the proximal demonstrative that makes reference to an entity near the deictic centre, and *kt* is the distal one that makes reference to an entity distant from the deictic centre. When these demonstratives are used with a locational function they will often take locational semantic case markers (locative, allative, perlative). Examples 20b and 21b show the demonstratives with the locative case, which have different forms when attached to common nouns (=*ano* for demonstratives vs. =*n* for common nouns):

(20) a. malind a dardar bä **yna** ta/kmarwn,
tortoise.ABS CONJ butterfly.ABS 3ABS **DEM.prox** 3du.U:rmpst/stay
gihe.
right.here
The tortoise and the butterfly, they lived **here**, right here.
WSEK1-G20170724-01FY10Farthm, 00:00:11.810 - 00:00:13.215

b. ynd bérbéru-va n\ovar/yn **yna=no** .

1ABS fear-COM 1sg.A:pfv.pret/arrive **DEM.prox=LOC**

I was scared when I arrived **here**.

WSEK1-B20150909-MissingMoneyMZ, 07:43.037 - 00:07:44.933

(21) a. nu qéki iy\avo/n. Nu **kt**

water container.ABS 1sg.A>3sg.U:pfv.pst/check water **DEM.dist**

yao bva ym.

NEG just 3sg.U:nphd/be

I checked the water container. There was no water **there**.

WSEK1-G20151001-13ThisMorningAN, 00:01:23.989 - 00:01:29.441

b. ämb evh=on n-ng\itro/yn **kt=ano** Yv

some day=LOC 1sg.A:pfv.pret:and/move **DEM.dist=LOC** place.name

On another day I moved to **there**, to Yv.

WSEK1-B20150804-01SnakeBiteTY, 00:02:07.901 - 00:02:10.333

Finally *yna* functions adnominally as a modifiers of head nouns (example 22a). Distal *kt* is difficult to distinguish as either adnominal or locational in uses. In example 22b, *kt* is preceding the word *dingi* 'dinghy', which can be taken as modifying it (i.e. 'that dinghy'), or it may be analysed as locational (i.e. 'there, on the dinghy').

(22) a. yndo [**yna** dram]$_{NP}$ y\roh/an.

1sg.ERG **DEM.prox** drum.ABS 1sg.A>3sg.U:ipfv.nphd/push

I am pushing **this** drum.

WSEK1-A20150917-1SinkingBA, 00:07:12.738 - 00:07:14.520

b. Ymo ynd [**kt** dingi=yan]$_{NP}$ gea w-ng\avli/,

3sg.ERG 1ABS [**DEM.dist** dinghy=LOC] when 3sg.A>1sg.U:pfv.past/put.in

q-ng\avli/nga kt=ano Boigu mutu-mutu=at.

3sg.A>1sg.U:pfv.fut:and/put.in DEM.dist=LOC Boigu island-RDP=ALL

When he put me inside **that** dinghy, he took me away to that place, Boigu Island.

WSEK1-B20150804-01SnakeBiteTY, 00:14:48.741 - - 00:14:54.040

Nmbo also has two *manner demonstratives*: *dena*, and *mna*. *Dena* is often translated as "like this", while *mna* is translated as "because". I have tentatively glossed these as 'f

Dena has multiple functions. It can refer back to past events or entities (example 23a), but also to draw attention to the immediate discourse (b). *Dena* also functions as a quotative immediately preceding or following a predicate of speaking, and preceding the direct speech (c).

(23) [Examples of dena]

a. [Backwards looking *dena*]

K-ng\ra/wm, ball t-n\kavo/ngaym. Ynd

1du.A:ipfv.rmpst:and/play ball.ABS 1nsg.A:ipfv.rmps:vent/chase 1ABS

ymo=va k-ng\ambo/nawm **dena**... Mära,

3sg.ERG=COM 1nsg.A:ipfv.rmpst:and/chase **DEM.mn** Mära

Martha, Waniko, Yumär. Ymova k\ambo/nawn.

Martha Waniko Yumär 3nsg.ERG=COM 1sg.A:ipfv.rmpst/play

We used to play, we used to kick the ball... We played with them **like that**... with Mära, Martha, Waniko, Yumär. With them I played.

WSEK1-G20170617-01Alqi02MQ, 00:04:13.294 - 00:05:16.997

b. [Drawing attention: Äugi is giving a demonstration of how to dig a

hole in the ground in order to plant tubers in it.]

Ok,	q\ake/ta!	**Dena**	däv	t\avngo/nge,
DISC	2sg.A>1sg.U:ipfv.imp/watch	**DEM.mn**	when	2sg.A>3sg.U:pfv.fut/do
gs.				
DEM				

Ok you watch me! **It's like this** when you will dig, here.

WSEK1-B20150720-01GardeningLM, 00:00:11.145 - 00:00:13.001

c. [Quotative *dena*: Doado is reporting the speech of Elma.]

Elma-m	w\mndo/i	**dena**	"wa	n\ango/ta."
Elma-ERG	3sg.A>1sg.U:pfv.pret/tell	**QUOT**	must	2sg.A:ipfv.nphd/return

Elma told me **like this**, "You must return."

WSEK1-B20150819-02DaruDS, 00:03:29.455 - 00:03:31.680

The use of *mna* is a little more specific. It appears to be an anticipatory demonstrative, and indeed it can be found with case markers which are also found on the anticipated segment of discourse (example 24). *Mna* may best be thought of as a type of place-holder demonstrative to indicate that the speaker is preparing to talk about a particular thing, but the precise lexical item is to follow. I have tentatively glossed *mna* as a 'forward looking demonstrative of manner' in similar vein to what Wilson (1980) calls a "forward referring" manner demonstratives for Ambulas (Ndu, Ambulas-Hanga-Hundi; PNG)(p.57).

(24) [Examples of forward referring *mna*]

a. Säläme k-ng\uvträr/wt, k-ng\ovyo/wt... **mna=yan**,
shirt.ABS 3du.A:ipfv.rmpst/take.off 3du.A:ipfv.rmpst/hang **DEM.fw=LOC**
wén kavwe=**an**.
tree branch=**LOC**
The shirts were taken off and hung... **so there**, **on** the tree branch
WSEK1-B20150804-01SnakeBiteTY, 00:14:48.741 - - 00:14:54.040

b. bä toge-toge tam bä de n-ng-owano/st
3ABS child-RDP.ABS 3nsg.U:ypst/be 3ABS CMPL 3nsg.A:pfv.pst/travel
ärgv=at, **mna=t** ei... wagiv=**ot**.
creek=ALL DEM.fw=ALL DISC fish=**ALL**
The children went to the river, for that, **you know**, **for** fish.
WSEK1-B20170614-03Smako02MQhm, 00:00:36.107 - 00:00:41.375

Additional comments on the origin of the manner demonstratives are in the appendix (A.5).

4.2.1.4 Adjectives

Following Siegel (2014a) and his analysis of Nama, Nmbo is analysed as having a semantic class of adjectives. Adjectives attribute some quality or characteristic to a noun (example 25a), and can also be used predicatively with a copula (25b). Adjectives are also homophonous with semantically related nouns. For example *wavta* (adj. 'hot') has a heterosemous word *wavta* (n. 'heat'), which can function as a head noun taking case marking (26a), and can also be modified by another adjective (26b).

(25) [Some common expressions.]

a. Bm [**mer** mrz] n/m.
2ABS good daughter 2sg:nphd/be
You are a good daughter.

b. Evh **wavta** y/m.

sun.ABS hot 3sg:nphd/be

The sun is hot.

(26) [*Wavta* as a noun.]

a. [Taqm is telling a story about a demon which was vanquished by a raging fire.]

Wavta-yam ge yaneyng niñi.

Heat-ERG FOC 3sgA>3sgU.pfv.pret/catch

The heat got the demon.

WSEK1-B20150805-01NinyiTS, 00:15:34.668 - 00:15:36.422

b. [**Wavta** kitong]$_{NP}$ ge nilawoi.

Heat large FOC 3sgA.pfv.pret/enter

The great heat went through [the tree].

WSEK1-B20150805-01NinyiTS, 00:15:34.668 - 00:15:36.422

The precise number of words that constitute the adjective class is currently unknown. Döhler (2018) notes what separates property nouns from adjectives in Komzo is the latter's inability to enter a possessive construction (p.88). I have not systematically tested these for Nmbo. The following list of Nmbo adjectives is based on proposed semantic types proposed by Dixon and Aikhenvald (2004) that are common for closed class adjectives.

- Dimension: *nmbn, nmbnarvi* 'small' ; *kitong* 'large' ; *tomba* 'long', *wra-wra* 'thin'
- Physical Properties: *bdrät* 'wet' ; *wao* 'ripe' ; *hrnat* 'dry'; *krkv* 'rotten/old'
- Age : *totr* 'new'
- Colour (Often synonymous with another natural feature like light or darkness): *site* 'white'; *frkta* 'red' ; *qèvte*, 'black'

- Human propensity: *nq* ‘anger’ ; *mwyat* ‘thought’
- Value: *mer* ‘good’; *mämä* ‘bad’

4.2.1.5 Quantifiers and Numerals

Quantifiers and numerals are a semantically unified class of words that function to specify the size of a set; the difference between the two is the degree of numerical specificity (von Mengden 2010:12). Nmbo quantifiers and numerals both behave similarly in terms of their semantics and morphosyntax. They both modify head nouns, which includes common nouns (example 27a) and personal pronouns (27b). Quantifiers and numerals occur immediately to the left or right of the head, but have not been tested for floating. Quantifiers and numerals can themselves function as pronominal heads of a noun phrase. Both can be reduplicated and take the dative *-e*, but these may also be fossilised collocations with idiosyncratic meanings. There is no distinction in form between the concepts of ‘all’ and ‘every’.

(27) a. **mñutyu** nne ge t\van/end.
all food DEM 3du.A>3sg.U:pfv.proxin/spoil
Those two have spoiled all the food.
WSEK1-G20170624-07RottenMNhm, 00:01:27.318 - 00:01:30.929

b. ynd **mñutyu** k\äuhe/tawm, .
1ABS **all** 1pl:ipfv.rmpst/be.happy
We were all happy.
WSEK1-B20150928-MPVisitLS, 00:04:55.421 - 00:04:57.030

Quantifiers tend to occur to the left or right of the head noun they modify. They can take case marking when they are the right-most constituent of the noun phrase core (example28a). This is in contrast to the sister language of Nen which does not do so (Evans 2017b:572). As mentioned, quantifier floating was not tested for, and there are no clear cases in the corpus. There are a handful of instances where quantifiers occur after a verb (example28b-c), but these are analysed as an afterthought rather than a float, as evidenced by the new intonation unit (orthographically expressed by ‘ , ’).

(28) a. Yve t\ävätu/tawt, [bisi **tärvär**-e].

3nsg.DAT 3pl.A>3sg.U:ipfv.rmpst/shoot [arrow **many**-DAT]$_{NP}$

They shot [the pig] multiple times, with many arrows.

WSEK1-G20150902-03PigGS, 00:12:22.160 - 00:12:26.700

b. Yao, wrseh-är är kt y/mengaran, **tärvär**.

NEG run-AGN.NMZR people DEM.dist 3pl+:nphd/live **many**

No, many of them are migrants living there.

WSEK1-A20150923-07TzZiBA, 00:05:58.640 - 00:06:00.555

c. Mrz-mrz-mn wah t-ng\räv/nawn, **tärvär**.

girl-girl-ORI song 1sg.A>3pl+.U:ipfv.rmpst:make **many**

I was composing lots of songs about girls.

WSEK1-B20170627-03SmakoSL03MQ, 00:00:06.335 - 00:00:06.335

The quantifier *ämb* has been observed in reduplicated form, taking an *-e* suffix. The compositional semantics of this reduplicated form are not entirely clear, and this *ämb-ämbe* may actually be a calcified collocation. When translating, Nmbo speakers will give *ämb-ämbe* as "for good" (and it is listed as such in the current form of the dictionary, examples 29)

(29) a. [Non-reduplicated use of *ämb*]

Gundma zi g=ym, **ämb** got-got=ro ä

Gunduma language.ABS TOPIC **some** word-RDP=RST 3ABS

n "ayäre/tan.

1sg.A:ipfv.nphd/hear

The Gunduma language [TOPIC] I hear only **some** of their words.

G20170803-04RichardRG02MQ5s, 00:09:46.990 - 00:09:48.900

b. [Bunai is relaying the thoughts he had when the dinghy he was on board was not being manned well, and he was concerned about where they will end up.]

de	**ämb-ämb-e**	bä	ynngm	ei.
ALR	**some-RDP-DAT**	3ABS	1nsg.U:nphd/go	DISC

We might go all the way out [into the ocean] **for good** [and never return].

WSEK1-B20150928-MPVisitLS, 00:04:55.421 - 00:04:57.030

Nmbo has a traditional numeral system functioning alongside an English numeral system. The English system functions almost exclusively as cardinal numerals. They function as lexical noun for dates, and years. Neither the English nor traditional system make formal distinctions between cardinal and ordinal numerals.

The traditional numeral system is a typologically rare senary (base six) system which is found throughout SNG (Donohue 2008, Hammarström 2009, Evans 2009). The basic values from one through to six constitute the building blocks for values greater than six. The basic values of the senary system are relatively frequently used, though it is also common for English numerals to be used in their stead. Speakers are in strong agreement about the form from one to six, and there is no variation. Upwards of six, however, speakers tend to use English numerals. When the senary system forms are used, there is variation. Table 4.14 shows the results of one elicitation session, with both Kerake and Yarne varieties represented.

The senary system of higher values are numerals for ascending powers of six. Evans (2009) hypothesises that the emergence of this particular numeral system was due to the cultural context of yam counting. Yam counting is a ritualised activity that has two men carrying three yams each to form a pile, and this pile is made up of the six yams deliberately laid out in an aesthetically pleasing manner[6]. To this day yams are often bundled into groups of six (figure 4.14).

Previous work by Martin (2001) shows a term *wi* which supposedly means 6^6, but this was not something that my consultants gave. I was given the word *mämä wim* for 6^6 (which literally translates to "bad scent"). Martin has 6^7 as *mämä wim*.

[6]See Evans (2009) for a contemporary account in the neighbouring Bimadbn village, and also Williams 1936: 225-227 for an older account

Value	Base 6	Kerake	Yarne
1	1	*ämbru*	*ämbro*
2	2	*sombwi*	*sombio*
3	3	*nambi*	*nambio*
4	4	*somba sombwi*	*sombio-sombio*
5	5	*widma tndro*	*widma tndro*
6	6	*ämbru for*	*ämbro fur*
7	6+1	*ämbru for mé ämbru*	*ämbro fur ämbro*
8	6+2	*ämbru for sombwi*	*ämbro fur ämbro sombio*
9	6+3	*ämbru for nambi*	*ämbro fur ämbro nambio*
10	6+4	*sombwi tnd brobro*	*ämbro fur ämbro sombio-sombio*
12	6X2	*sombwi for vor*	*sombio furvur*
36	6^2	*frta/ämbru frta*	unknown
216	6^3	*taromba*	unknown
1,296	6^4	*damno*	unknown
7,776	6^5	*wärämka* /wæ.ræ.mɐ.ka/	unknown
46,656	6^6	*mämä wim**	unknown

Table 4.14: Nmbo senary numeral system 1-12. * = Martin (2001) gives as *wi*.

Figure 4.14: Yam counting in Bevdvn, June 2015.

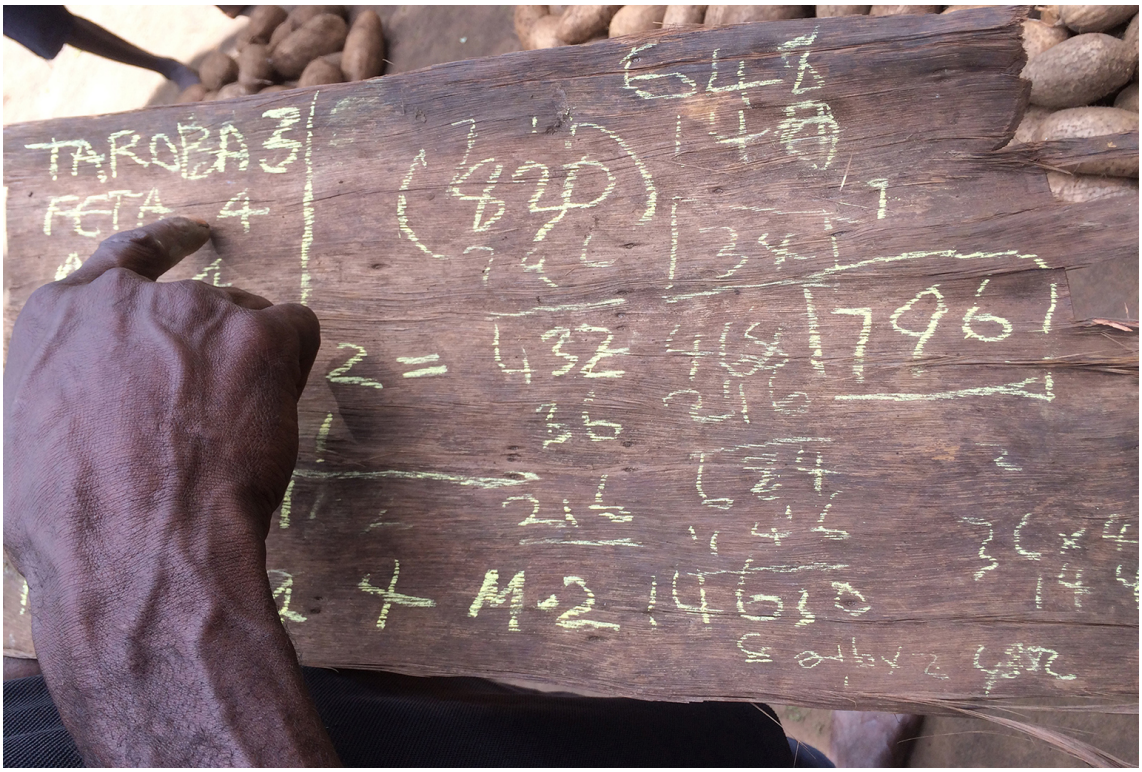

Figure 4.15: Yam counting in Bevdvn, June 2015. Writing by Bunai Äniba.

One point of interest is the form of the 6^2 value. The form *frta* listed in table 4.14 is an elicited one, but I have observed the use of the form *feta* during the course of my fieldwork. In Williams (1936), the 6^2 form is listed as *peta* (p.227). This suggests that the form *frta* is perhaps a recent innovation, with the form *feta* as the intermediary form (figure 4.15). Note that the cognate for *feta* in Nen is *peta*.

Numerals can be used in a non-numeric sense, i.e. the word *ämbru* 'one' being used in the sense of 'small thing' or 'a bit of something'. It may be possible for numerals used non-numerically to take a restricted set of case markers, but this has not been tested. The example given below with the dative (example 31) may in fact be a lexified collocation.

(30) a.

Dena	ämb	**ämbra**-e	ädi	ym.
DEM.mn	some	**one**-DAT	EMPH	3sg.U:nphd/be

I know bits and bobs.

WSEK1-G20150902-07MarriageSY2, 00:07:45.645 - 00:07:47.316

Numerals can, like quantifiers, be reduplicated and followed by a dative *-e*. The reduplicated numerals are akin to what Evans (2017b) describes as *distributive numerals* in Nen. Nen forms manner adverbs with the suffix *-ae* on numerals to produce mean '*n* each', '*n* by *n*', or 'per *n*' (Evans 2017b:590). The dative in Nmbo is also used to form manner adverbs (section 4.4.4.2), so a similar phenomenon likely occurs in Nmbo.

(31) a.

dena	ge	ta-ng/maro	**ämbru-ämbru-e**
DEM.mn	DEM	3pl.U:rmpst/go	**one-RDP-DAT**

Like this, they went, one by one.

WSEK1-B20170627-03SmakoSL03MQ, 00:04:09.695 - 00:04:11.555

b. [Terav is telling a story about the time he was bitten by a Papuan black snake, and how his vision doubled, and how he saw two of everything.]

ta\mnd/en	“bm	ge	e/ren,	**sombwi-sombwi-e**
1sg.A>3du.U:ipfv.ypst/say	2ABS	DEM	2sg.U:nphd/be	**two-RDP-DAT**
e/ren e.”				
2sg.U:nphd/be DISC				

I said to them “You both are two each.”

WSEK1-B20150804-01SnakeBiteTY, 00:07:43.583 - 00:07:46.873

4.2.1.6 Interrogatives

Interrogatives in Nmbo are question words that indicate the speaker's lack of knowledge pertaining to the complete identity of a referent. Nmbo has interrogative words for human and non-human nouns, locations, temporals, numerals, reason, and manner (table 4.15).

Nmbo, like its sister language Nama, can use interrogatives as relativisers (section 4.9.7); a cross-linguistic rarity according to Siegel (accepted).

A couple of the locational interrogatives appear to be formed by binding locational clitics to the base form, and it is unclear whether these are produced productively with the full set of spatial demonstrative clitics, or if these are lexicalised occurrences. It is likely that there are other such locational interrogatives formed by morphology, which are not shown on this list.

There is one interrogative for animates in Nmbo with three case forms: *eve* (absolutive), *emo* (ergative singular), and *emovem* (ergative non-singular). These follow the case marking paradigm for human nouns, distinguishing between an absolutive and ergative form, as well as having the ergative form make a singular vs non-singular distinction. The absolutive form *eve* is number invariant and is indexed on monovalent verbs as a third singular (example 33). The ergative *emo* (singular) and *emovem* (non-singular) trigger agreement on divalent verbs, also as a third singular.

(32) a. **eve** **y**/m?
who.ABS **3sg.U**:nphd/be
Who is it?

Type	Form	Meaning
Human	*eve*	Who is? (who.ABS)
	emo	Who did X? (who.ERG)
	emovem	Who (nsg) did X? (who.ERG.nsg)
	ende	Whose? (who.POSS)
Non-Human	*nmbo*	What?
Locational	*ka*	Where?
		Anywhere, somewhere
Locational Goal	*kano*	Where to?
Locational Source	*kanmn*	Where from?
	kanma	Which one?
Temporal	*sam*	When?
Numeral	*snamb*	How many?
Reason	*na*	What reason?
	non	Why do?
Manner	*dr*	What happened?
	dran(d)ma	How?

Table 4.15: Interrogatives.

b. **eve** n-owav/**t**?

who.ABS **3sg.A**:ipfv.nphd/speak

Who is speaking?

Fieldnotes 2015 Book 3

(33) a. **emo** t\avro/**tao**?

who.ERG **3sg.A**>3sg.U:ipfv.rmpst/do

Who did it?

b. **emovem** t\avro/**tawt**?

who.ERG.nsg **3nsg.**A>3sg.U:ipfv.nphd/speak

Who is speaking to him?

Fieldnotes 2015 Book 3

As mentioned, interrogative pronouns can function as relative pronouns to introduce relative clauses (section 4.9.7). The relative pronoun agrees in number with the finite verb in the main clause; this is marked on the a-suffix of the verb, which indexes the person and number of the actor argument.

(34) a. yna är-m **emo** wñe-h zi ta

DEM.prox man.ERG **who.ERG.sg** lie-INF story.ABS 1SG.DAT

q\awaito/**t**.

3sg.A>3sg.U:ipfv.ypst/tell

This is the man who told me lies.

b. yna är-m **emovem** mamwi t\n-rs/**awt**.

DEM.prox man-ERG **who.ERG.nsg** pig.ABS **3nsg.A**>3sg:ipfv.rmpst:ven/carry

These are the men who were carrying the pig.

Fieldnotes 2017 Book 1: 75

The possessive *ende* seeks information on the identify of a possessee. It is equivalent to the English 'whose'. Note that the form is homophonous with the possessive suffix (section 4.4.4.3).

(35) a. **ende** qéki **y**/m?

whose container.ABS **3sg.U**:nphd/be

Whose container is it?

Fieldnotes 2015 Book 3

The object interrogative pronoun *nmbo* 'what' is used to seek specification of inanimates (example 36a) or intent (b). *Nmbo* is also the source of the language

name of the Kerake people, and the sister languages of Nmbo also use their respective object interrogatives as the language name. The object interrogative *nmbo* has been observed to occur with the comitative clitic (c).

(36) a. **Nmbo** y/m?

what **3sg.U**:nphd/be

What is this?

Fieldnotes 2017 Book 3

b. bmo **nmbo** bä y\avro/t?

2sg.ERG **what** FUT 3sg.A>3sg.U:ipfv.nphd/do

What will you do with that??

WSEK1-G20151015-06BurnedMN, 00:00:50.861 - 00:00:51.935

c. konga ämb di y\avro/tat?... **nmbo=va**?

cassava.bake some ALR 3pl.A>3sg.U:ipfv.nphd/do **what=COM**

Did they make some cassava bake? With what?

WSEK1-G20151015-06BurnedMN, 00:09:32.699 - 00:09:37.244

Locational *ka* asks about the identity of a location. It forms part of the demonstrative paradigm of case marking. *Ka* has also been observed functioning like a spatial indefinite pronoun (example 37b).

(37) a. **ka** y/m?

where 3sg:nphd/be

Where is it?'

Fieldnotes 2015 Book 3

b. [Richard is telling a story about how he suspects a spirit attacked him the night before.]

nonavi	yna	**ka**	ge	y/akiyongr	ge
maybe	DEM.prox	**somewhere**	DEM	3sg.U:nphd/stand	DEM

y/m?

3sg.U:nphd/be

It [i.e. the spirit] might be standing somewhere here?

WSEK1-G20151013-10SpiritRG, 00:03:50.970 - 00:03:53.021

The temporal *sam* is concerned with seeking information about the time of an event. *Sam* has not been observed occurring with any case markers. It can also function as a relativiser (example 38b).

(38) a.

sam	fivi	y\ivo/i?
when	INTS	2sg.A>3sg.U:pfv.pret/finish

When did you finish it?

WSEK1-A20150924-12BasketKA, 00:02:36.870 - 00:02:39.921

b.

ynd	kitong	**sam**	n\amdo/yn...	ynd	awatambna-h
1ABS	big	**when**	1sg:pfv.pret/become	1ABS	learn-INF

n\ovngo/yn...	yäme	rama-h...
1sg.U:pfv.pret/begin	mat.ABS	make-INF

When I became bigger... I started learning... mat-making

WSEK1-G20170713-02MQMW, 00:06:26.448 - 00:06:35.385

Numeral *snamb* [sɐ.namb] is best translated as 'how many; how much', and has not been observed occurring with any case markers. The distinction between count and mass nouns is not made in Nmbo.

(39) a.

Snamb	toge	n/m?
how.many	child.ABS	2sg.U:nphd/be

How many children do you have?

WSEK1-G20141008-02CoconutEN, 00:04:07.579 - 00:04:09.160

b. mani **snamb** t/maro?

money.ABS **how.many** 3sg.U:rmpst/be

How much money was there?

WSEK1-B20150909-MissingMoneyMZ, 00:07:20.825 - 00:07:22.083

There are three different forms which are translated as 'why', inquiring into the manner, reason, and cause of some person's action: *Non*, *na*, and *dr*.

Non is perhaps best translated as 'what reason?'.

(40) a. **non** n\e/nd?

why 2sg.A:pfv.pst/cry

Why do you cry?

Fieldnotes 2015 Book 1:

b. mamwi gs=ym... **non** herge y/m?

pig.ABS [TOPIC] **why** life 3sg.U:nphd/be

This pig... why is it alive?

WSEK1-B20141007-PigYS, 00:15:03.555 - 00:15:04.773

c. [Jacklyn is explaining how the gänzän spirit placed the Nen speaking village down by itself, explaining how Bimadbn is the sole Nen-speaking village in the area. She is asking the interviewee how this situation came to be.]

Bimadvn y\kmo/i yämbia [...] **non**

Bimadvn 3sg.A>3sg.U:pfv.pret/lay.down 3sg.ALN **why**

y\kmo/i dena?

3sg.A>3sg.U:pfv.pret/lay.down DEM.mn

Bimadvn was placed itself... Why was it placed so?

WSEK1-B20150813-03PerceptualHZ, 00:06:21.500 - 00:06:26.500

Na is also typically translated as 'why', but in some instances, the translation offered is closer to the interrogative object pronoun 'what'. *Na* can perhaps be analysed as some modifier within an interrogative phrase (example41b and (c)).

(41) a. bm **na** kngmaro, bende ama-nde hamba=t?
2ABS **why** 2sg.U:rmpst/go 2sg.POSS mother-POSS village=ALL
Why did you go to your mother's village?
WSEK1-G20170617-01Alqi02MQ, 00:06:11.284 - 00:06:14.384

b. **na** kembone k-ng\ambn/awte?
what game 2pl+.A:ipfv.rmpst/play
What games did you all play?
WSEK1-G20170617-01Alqi02MQ, 00:03:24.069 - 00:03:25.779

c. kor yao mywat w/m, **na** yz=n,
again NEG knowledge 1sg.U:nphd/be **what** year=LOC
tande ama, n\gla/mng...
1sg.POSS mother 3sg.A:pfv.pst/born
I don't really know in what year my mother was born...
WSEK1-G20170713-02MQMW, 00:01:05.888 - 00:01:09.975

Dr is concerned with the identity or manner of an action, so could be best translated as 'how?'. *Dr*, unlike *non*, appears to be part of a paradigm with demonstrative case markers, with forms such as *dr=andmae* 'how to do?' (example 43).

(42) [*dr* 'how']

a. [Wäwi recounts how she spotted a pig in the past, and reports the thought that came across her mind.]

dr ä y\avro/tan?

how FUT 1sg.A>3sg.U:ipfv.nphd/do

What am I going to do to it?

WSEK1-B20150924-HuntingWZ, 00:00:29.369 - 00:00:30.352

b. [Yavs is imagining what his descendent in the future will say when they hear the recording we have just made.]

o **dr** tvende deve de k\owv/tao

EXCL **how** 1nsg.POSS father.ABS ALR 3sg.A:ipgv.rmpst/speak

yna zi.

DEM.prox story.ABS

Oh yes that's how our father told that story.

WSEK1-B20151018-03BevYZ, 00:01:27.086 - 00:01:29.291

(43) [*drand(m)ae* 'how / what do?']

a. **dr=andmae** bä y-ng\avro/t?

how FUT 2sg.A>3sg.U:ipfv.nphd:and/do

What will you do? [lit. How will you do it?]

WSEK1-G20151015-06BurnedMN, 00:07:46.896 - 00:07:48.284

b. yna snar wagiv **dr=anmae** y\avro/tat?

this fish.type fish how 2sg.A>3sg.U:ipfv.nphd/do

This small fish, what do you do with it? [lit. How do you do this fish?]

WSEK1-G20151015-06BurnedMN, 00:09:23.808 - 00:09:25.978

4.2.1.7 Indefinite Pronouns

A small set of indefinite pronouns that end in the form *vi* have been identified for Nmbo. The most typical occurrence of this suffix is following the human

interrogative pronoun, e.g. who.ERG + vi → *emovi* 'anyone'.[7].

Form	Meaning	Semantic Case?
emovi	Anyone, someone	unknown
nmbovi	Something	unknown
kanmavi	Anyway, someway	unknown
nonavi		unknown

Table 4.16: Indefinite pronouns.

I have only been able to elicit four forms following this pattern: *emovi* 'someone, anyone', *nmbovi* 'something, anything', *kanmavi* 'someway, anyway', and *nonavi* 'maybe, perhaps' (table 4.16). The *vi* may possibly be a semi-productive morpheme, and has been observed occurring at the right edge of a noun phrase containing an interrogative (example 44). It is possible that this example of right-edged *vi* is in fact an instance of phonetically reduced *fivi*. This would be an interesting case of variation to investigate as a process of ongoing grammaticalisation.

(44) a. [**na rokar**]=**vi** ämb yao y\avaro/taym.
what thing=vi some NEG 1sg.A>3sg.U:pfv:pret+ta/think
We couldn't think of **anything**.

Nmbo Dictionary 2018 entry for *emovi*

Indefinite human *emovi* means 'somebody' or 'anybody'. *Emovi* is indexed on the verb by the a-suffix; in example 45a as the second singular, and in (b) as the third singular form. *Emovi* can have negative indefinite meaning 'nobody' in negated contexts (45b).

(45) a. **emovi** deu y\avro/**ta**.
anybody can 2sg.A>3sg.U:ipfv.nphd/dp
Anybody can do it.

WSEK1-G20170617-01Alqi02MQ, 00:06:11.284 - 00:06:14.384

[7]A possible origin for this is a fusion of the reduced form *fivi* 'body', 'real'. This would then fit a cross-linguistically observed pattern of indefinite pronouns being formed by combining some expression of definiteness to indefinite pronouns (Haspelmath 1997:194-196)

b. **emovi** yao ä t-ng\ane/**nga**.

anybody NEG FUT 3sg.A>3sg.U:pfv.fut/take

Nobody will take it. [lit. anybody will not take it]

Nmbo Dictionary 2018 entry for *emovi*

Indefinite inanimate *nmbovi* translates to 'something' or 'anything'. *Nmbovi* is indexed on the verb by the u-suffix in the third singular form.

(46) a. ymo **nmbovi** **y**\ros/

3sg.ERG **something** 3sg.A>**3sg.U**:pfv.past/search

He searched for something.

Nmbo Dictionary 2018 entry for *nmbovi*

b. ämb **nmbovi** yao **t**\avro/tao kt.

some **anything** NEG 3sg.A>**3sg.U**:ipfv.rmpst/do DEM.dist

She couldn't do anything there.

WSEK1-B20150805-01NinyiTS, 00:15:32.323 - 00:15:34.668

The indefinite manner pronoun takes the form *kanmavi*. It is a rare form that was elicited, and does not occur naturally in the current state of the corpus.

(47) a. **kanmavi** deu n-ng\m

any.way can 2sg.U:nphd/go

In any way that you can.

Nmbo Dictionary 2018 entry for *kanmavi*

Nonavi is consistently translated as 'maybe' or 'perhaps', functioning as a type of epistemic modal marker.

(48) a. **nonavi** tande mani ge tnane?
perhaps 1sg.POSS money DEM 2sg.A>3sg.U:pfv.proxin/get
Perhaps you got my money?

WSEK1-B20150909-MissingMoneyMZ, 00:06:09.706 - 00:06:11.080

b. wén **nonavi** ge n\wiho/t kkv
tree.ABS **perhaps** DEM 3sg.A:ipfv.nphd/sprout garden.ABS
kndn
inside
A tree may sprout in the garden.

WSEK1-B20151009-07WeedingKS, 00:00:39.918 - 00:00:42.088

4.2.2 Verbs

Nmbo verbs are a clearly identifiable word class.

- Morphologically: Inflect for tense and aspect, and index the person and number of up to two core grammatical arguments, with options for morphologically indicating a third argument.
- Syntactically: Tend to occur clause finally, cannot be the head of an NP without non-finite morphology.
- Semantically: Expresses events, states, and change of states.

Nmbo has *infinitive verbs* formed by suffixing an infinitiviser *-h* to the root (e.g.*ano-h* 'to grow'). These non-finite verbs are found in a range of subordinate clauses (examples 49), and can take additional case marking such as the locative and allative (section 4.4.1).

(49) a. Szräzr **wavro-h** t-ng\avngo/n
house.post.ABS **do-INF** 1sg.A>3pl+:pfv.fut/do
I started making the house posts.

WSEK1-02DimbanKSae, 00:03:36.780 - 00:03:38.390

b. Elma-m w-n\arawe/tamng toge **wake-h**=t.

elma-ERG 3sg.A>1sg.U:ipfv.prim:ven/request child.ABS **watch-INF**=ALL

Elma requested me to watch the child.

WSEK1-B20150819-02DaruDS, 00:00:10.296 - 00:00:14.578

At their most complex, Nmbo verbs can:

- Index the person and number of up to two arguments;
- Encode directional information that can be semantically extended to non-spatial use;
- Mark valency adjustments;
- Inflect for one of eleven different types of TAM.

The following example was also presented in the introduction of the sketch grammar as an example of the verb expressing values in all slots. Note the glossing in this example follows a different convention in order to put a templatic perspective on the verb complex (see preamble, *Notes on reading the interlinear glossing*).

(50) a. k-ng-a-waramo-ta-ng-e-ay-m

1sg.U:β-AND-APP-give-IPFV-PFV.FUT-PFV.DU-PRET-1nsg.A

I should have given it away to you.

Fieldnotes 2017, Book 1:16

The details of the verb template are presented in the verbal morphology section (4.5).

Inflectional verbal TAM paradigms are built upon a distinction of verbs as *imperfective* or *perfective* in aspect (section 4.7. Imperfective verbs can simplistically be characterised as durative, and perfective verbs are momentaneous and punctual (Siegel 2014b). Many verbs in Nmbo can form both imperfective and perfective paradigms, but some can only form one. Prefixing verbs can only form an imperfective TAM paradigm.

4.2.3 Particles

Particles are a class of words that do not inflect. Semantically there are three types: TAM, discourse, and adverbial particles. Particles typically precede the main verb of a clause, but may also follow the verb. Other linguistic items may intervene between a particle and a verb. This last criterion is the main reason for classifying this class of words as particles rather than as adpositions or bound morphemes. Syntactically there does not appear to be any distinction between TAM and adverbial particles.

- Morphologically: Do not inflect, and do not take any semantic case marking.
- Syntactically: Free order, but often found preceding or following verbs.
- Semantically: Expresses TAM, discourse, and adverbial concepts.

TAM particles specify or add TAM dimensions to a verb. Some particles, such as the future marking *bä* seem to occur only with imperfective verbs, and are the only means to specify future tense for these verbs. Others such as *bva* or *dva* seem to specify completion of an event.

Examples and further description of the particles are presented in section 4.3.

4.2.4 Markers and typical collocations of discourse

Here I have listed a handful of typically collocations and discourse particles found in Nmbo.

Discourse marker *e~*, signals the passing of time, and is translated as “until". The vowel is held anywhere up to 1 second in duration.

ädi: an intensifier, a particle of emphasis, akin to “truly”, “really”.

ei: an exclamation, akin to “hey” in English.

(bä) bädi: an intensifier, a particle of emphasis, akin to “truly”, “really”.

foyo ~fyo: translated as “that’s fine”, often used to indicate a change in discourse topic, translatable as “and then”.

kor yao: “not really”. Same as *yao kor*.

mämä hrae: “in a bad way", used in instances to express stress and anxiety. Conversely, *mer hrae* is “in a good way".

se: an intensifier, a particle of emphasis, akin to "truly", "really".

woi dena yamoyo: "again"

yao kor: "not really". Same as *kor yao*.

4.3 Particles

This section goes through all the currently known temporal and modal particles of Nmbo. Discourse particles are shows in section 4.2.4. For the definition of Nmbo particles as a word class, see section 4.2.3. A summary of types and glosses are shown in table 4.17

Form	Function	ILG	Translation	Section
Temporal				
bä	temporal: future	FUT	will	4.3.1.1
de ~ di	temporal: iamitive	ALR	already	4.3.1.2
bva, dva	temporal: completive	just	already, nearlly	4.3.1.3
mé ~me ~ mi ~$m^w i$	temporal: continuative; modal: epistemic	CONT	still, will	4.3.1.4
dumen	temporal: always, habitual	HAB	always	4.3.1.5
Modal				
deu	modal: deontic	can	can	4.3.2.1
wé ~wa	modal: deontic	must	must	4.3.2.2
mato	modal: epistemic,	perhaps	perhaps	4.3.2.3
gea	modal: epistemic: hypothetical, supposition,	SUPP	when, if	4.3.2.4
yta	modal: frustrative	FRUS	try to no avail	4.3.2.5
Negation				
yao ~yawa	negation of nominals and predicates	NEG have not	not, do not,	4.3.3.1
Demonstrative				
ge, *gs*	manner	DEM	right t/here, like so	4.3.3.2

Table 4.17: Nmbo temporal and modal particles.

4.3.1 Temporal Particles

4.3.1.1 Bä (future)

Bä is the future marking TAM particle. It occurs with the imperfective non-prehodiernal tense. It is homonymous with the third person absolutive pronoun, which sometimes makes it difficult to tell whether what appears in a clause is the particle or pronoun. This future marking particle also often appears in the form *ä* without the consonant. The third person absolutive pronoun, also *bä*, does not occur in the form *ä*, although I have not quantified this (example 51b).

(51) a.

yndon	bä	y\aito/tan	zi,	gnoso
1sg.ERG	FUT	1sgA>3sgU:ipfv.nphd/tell	story.ABS	now

yägu-mn.

morning-ORI

I will tell a story about this morning.

WSEK1-G20151001-13ThisMorningAN, 00:00:22.092 - -00:00:28.835

[Mwigm is telling a story about how his pineapple garden got burned down by a bush fire. Note that the second future particle is in the form *ä*.]

b.

Totr	ynd	zäru	**bä**	w/mn...	zäru	kitong
now	1ABS	hunger	**FUT**	1sg.U:nphd/be	hunger	big

ä	k\ovar/nga!
FUT	3sg.A:pfv.fut/arrive

Now I will be hungry... a great hunger will come!

WSEK1-G20151015-06BurnedMN, 00:01:36.200 - 00:01:39.800

4.3.1.2 De ~ di ('already', iamitive)

De, or sometimes *di*, is translated varyingly as 'already' or 'just', suggesting that it is a kind of completive particle that signifies the completion of an event in terms of discourse. I follow Döhler (2018) and adopt the term 'iamitive' (from

the Latin *iam* 'already', Olsson 2013) and gloss it by the label ALR 'already'. This particle occurs with a variety of main verb TAM specifications.

(52) [The speaker is telling a story about the time he was nearly gored to death by a wild pig. The following utterance is him describing why he approached the wild boar after he had fired his arrows at it:]

a. K\e\mna/ngn **de** d-h t\amd/n
1sg.A:pfv.fut/think **ALR** shoot.INF 1sgA>3sg.U:pfv.proxin/apply.pressure
kr=ro=tio
death=RST=until

I thought that I shot it to death already [until it was completely dead]"

WSEK1-B20150727-01PastCustomMZ, 00:00:12.638 - 00:00:17.490

b. Nu kt t/maron. N\atn/amn **de**
water DEM.dist 3sg:rmpst/be 1sg.A:ipfv.prim/jump.over **ALR**
n\ouwi/yn ynd.
1sg.A:pfv.pret/fall 1ABS

There was water there. I jumped over and fell in [and stayed put].

WSEK1-B20150804-01SnakeBiteTY, 00:08:55.291 - 00:08:58.591

Noticeable is the occasional tendency for *de* to become phonologically merged with the prefix of the following main verb. If the prefix of the main verb is *y-* or *yn-* (third singular or first plural undergoer), the *de* particle leans in and often sounds like the prefix is *d-* or *dn-*.

(53) The speaker is talking about the time they were travelling to Daru by dingy.

a.	Darut	enemsam,	nu	bende	**dingosm**.
	Daru=t	e\nem/sam	nu	bende	**de**=y\ingo/sm
	Daru=ALL	1pl.A>2pl.U:pfv.pst/escort	liquid	2sg.POSS	ALR=3sgU-spot-1plA

We will take you all to Daru, we've already found your petrol.

4.3.1.3 Bva/dva ('just; nearly')

The precise difference between **bva** and **dva** is not entirely clear, but I propose it is a kind of immediate future or past completive. One may translate this as equivalent to the English "already" or "nearly".

Firstly, a quick comment on additional variant forms must be made. The first variant is the form **gva**, which is also translated as "just" or "already", and occurs in contexts that appear similar to those of *dva* and *bva*. Some of my consultants explained that this is merely a contracted form of *gs bva* or *gs dva*, where *gs* is a demonstrative particle (example 54b, section 4.3.3.2). The other variant is where the initial consonant has been dropped, resulting in a form akin to [ɐβɑ] (orthographically represented as *áva*)(example 54c). The existence of these variant forms suggest that the semantic distinction between *bva* and *dva* are perhaps not that great, and that these two forms are variants of a single linguistic variable.

Bva/dva provide a completive interpretation to events. There are a handful of examples, however, where it provides an interpretation of "nearly" (54c). This has also been attested for Nmbo's sister language Nama (Siegel, pers. comm.).

(54) a.

q\mndo/nga	de,	"o	bende	rusa	**dva**
3sg.A>1sg.U:pfv.fut/tell	QUOT	DISC	2sg.POSS	deer.ABS	**just**
y\gme/tam,		aha-m	**bva**	y-n\krmb/n!	
1pl.A>3sg.U:ipfv.nphd/strike		dog-ERG	just	3sg.A>3sg.U:pfv.pst:ven/chase	

He told me "Oh, we killed your deer, the dogs just got it."

WSEK1-B20150924-HuntingWZ, 00:05:05.532 - 00:05:08.366

[Jacklyn is doing in the Cassowary Picture Task, and is describing a picture in front of him which he has never seen before.]

b. orng=ama **gva** y-n/m äuya toge

jugle=PERL **gs=bva/dva** 3sg.U:ipfv/come cassowary small

jugle=PERL **just?** come cassowary small

y\ingo/

3sg.A>3sg.U:pfv.pst/spot

3sg.A>3sg.U:pfv.pst/spot

She's just coming through the jungle, and she spotted a cassowary chick.

WSEK1-B20170610-03JY03Casshm, 00:00:55.194 - 00:00:57.114

c. Kt g=ym **ava** nowyesn dena

there [TOPIC] **?** 1sg:pfv.pret/fall DEM.mn

There I almost fell over, like this.

WSEK1-G20151013-10SpiritRG, 00:02:44.291 - 00:02:45.971

4.3.1.4 Mé (continuative)

Mé has a wide range of phonetic realisations, ranging from [mɐ~mə] to [mɪ~m^{w}ɪ].

This particle performs two functions. One is a clear instance as a continuative particle. It typically precedes the verb that it modifies, expressing the concept 'continued doing V' (example 55b, the second instance). The expression *mé ym* with the continuitive preceding the copula is used to mean 'it is still the case' (example 55a).

The other function is *mé* as an epistemic modal marker translatable as 'should' (example 55b, the first instance). The semantic connection between continuative and epistemic modal is perhaps one of expectation and continuation of situation/circumstance.

(55) a. bä dena k\emn/ao malind **mé**

3ABS DEM.mn 3sg.A:ipfv.rmpst/think tortoise.ABS **CONT**

y/m.

3sg.U:nphd/be

She thought that the tortoise was still there.

WSEK1-G20170724-01FY10Farthm, 00:05:15.695 - 00:05:17.364

[Ḡño is telling a story about two anthropomorphised birds, kakma (hornbill) and kakayam (bird of paradise.) The two birds were frolicking in the water, but kakayam starts to get cold and wants to leave. The following example starts with a quote by kakayam:]

b. "Kakma **mé** k\urt/em." Kakma yände

hornbill **CONT** 1du.A:pfv.proxin/come.out hornbill.ABS 3sg.POSS

mende t/maro, anu **mé** ge k\umben/ao.

desire 3sg.U:irmpst/be bathing **CONT** DEM 3sg.A:ipfv.rmpst/bathe

"Hornbill, we should both come out [of the water]." [But] Hornbill wanted to keep bathing.

WSEK1-G20170723-01GwN10KK, 00:02:32.629 - 00:02:37.852

4.3.1.5 Dumen ('always', habitual)

Dumn ([du.mɐn]) is a habitual particle which is often translated as 'always', and occasionally as 'used to'. Note in examples 56a and (b) the translations given were 'always', but the translation 'used to' would also suffice, while example (c), the English 'always' would be odd-sounding.

(56) a. tande budär fivi gs t/maro Mwendak
1sg.POSS friend real DEM 3sg:ipfv.rmpst/be name
ynd ymo=va k\itro/wm **dumn**.
1ABS 3sg.ERG=COM 1du:ipfv.ypst/move **HAB**

My best friend Mwendak, the two of us would **always** travel around together. [lit. with her the two of us would wander around].

WSEK1-G20170720-01Gerida09Final, 00:00:35.431 - 00:00:38.986

b. [Bunai is explaining how the coastal village of Tais came to be established. In the following utterance he is talking about the movement of the Torres Strait Islanders (the Boigu Islanders) around the coast of southern New Guinea.]

Boigu=mn är g=ym **dumen** ta-n/maro,
Boigu=ORI people TOPIC **HAB** 3pl:ipfv.rmpst/come
kt=ano yna-ma.
DEM.dist=LOC DEM.prox-DEM.PERL

The Boigu people would **always** go through there [lit. come through there].

WSEK1-A20150923-07TzZiBA, 00:01:01.828 - 00:01:04.236

c. mamwi kt **dumen** n\ovro/t
pig.ABS DEM.dist **HAB** 3sg.A:ipfv.nphd/arrive

A pig used to arrive there.

WSEK1-G20151013-10SpiritRG, 00:01:13.621 - 00:01:15.451

4.3.2 Modal Particles

4.3.2.1 Deu (deontic 'can')

Deontic **deu** is very infrequently represented in the corpus. It is translated as "can", expressing the possibility of some event occurring. The sister lan-

guage Nama also has a word *teu*, which is translated as "can" (Siegel pers. comm.).

(57) a. Nes **deu** k\amdo/nga, yavav=etro däv,
nurse.ABS **can** 3sg.A:pfv.fut/become 3nsg.GOAL=RST when
mer gnogo=va gea y/m.
good head=COM SUPP 3sg:ipfv.nphd/be
She can become a nurse, if she is smart [lit. with a good head].

WSEK1-A20151001-09PaulaRM, 00:01:23.140 - 00:01:26.348

b. emovi **deu** y\avro/ta.
anyone **can** 2sg.A>3sg.U:ipfv.nphd/do
Anyone can do it.

Nmbo Dictionary Draft Version 2018, entry *deu*

4.3.2.2 Wé (deontic 'must', inevitably)

Wé or *wa* is a particle that prototypically expresses obligation, though there are some uses of this particle that suggest the usage is a little more flexible. It is most often translated as 'must', but I suggest that it is more generally expressing a sense of inevitability, both wishful and perceived.

The typical deontic uses can be interpreted as the speaker expressing or demanding an inevitable outcome.

(58) a. [Terav is telling a story about the time he was bitten by a Papuan Black snake. The following utterance is a relaying of his thoughts after he had told some people to go and get help for him.]

tande zi **wé** y-ng/m yve=vav.
1sg.POSS story **must** 3sg.U:ipfv.nphd/go 3nsg.DAT=GOAL
My message must go to them.

WSEK1-B20150804-01SnakeBiteTY, 00:12:23.650 - 00:12:25.896

There are some uses where the translation is more akin to "surely", indicating that the speaker believes, or believed, that some outcome or situation would be inevitable. In the example below the speaker is relaying his feelings of a time when he was camped in the bush and became frightened that a demon was near by. The original translation offered was " it might bite or kill me", which I have re-interpreted to mean "it will surely bite me", given the sense of inevitability that is entailed in the "must" uses of *wé*.

(59) a. ädi k\urtu/tan, ämb-ämb-e, krava **wé** w\rn/a,
INTS 1sg:ipfv.ypst/get.out some-RDP-DAT EXPL **mus**t 3sg.A>1sg.U:pfv.pst/bite
o, w\gme/t!
CONJ 3sg.A>1sg.U:ipfv.nphd/kill

I jumped out for good, oh dear, it **will surely** bite me, or kill me!

WSEK1-G20151013-10SpiritRG, 00:04:04.190 - 00:04:07.346

4.3.2.3 Mato 'perhaps; maybe'

Mato is often used as an offering of suggestion, translated as 'perhaps ; maybe', and conative 'try'. Unlike frustrative *yta* 'try', *mato* flags the speaker's intent to attempt an action, or a speaker's suggestion to another person to attempt another action. In Nen, the cognate particle *meta* is described as being used for polite requests, which I believe is also true of Nmbo.

(60) a. [Kawas is reporting what he said to his friends when they tried to help a sick man. He suggests to his friends that they touch his body to check the temperature.]

t\amdo/ngn, "**mato** bmovem t\ere/nge."
1sg.A>3pl.U:pfv.fut/tell **perhaps** 2pl.ERG 2pl.A>3sg.U:pfv.fut/hold

I told them, "Perhaps you all should touch him."

WSEK1-B20150817-02DimbanKSae, 00:09:53.000 - 00:09:54.935

b. [Zoga is looking at a page from the Cassowary Picture task. He is speculating as to what the two characters are saying to the crying

cassowary.]

Mato	dena	y\mi/tangeaend,		“bm	non
perhaps	DEM.mn	3du.A>3sg.U:pfv.pret+ipfv+fut/ask		2ABS	why
n\end/?”					
2sg.A:pfv.pst/cry					

Perhaps those two are asking [the cassowary] “Why do you cry?”

WSEK1-G20170706-03ZG04Retell, Approx. 00:03:34

4.3.2.4 Gea ‘if ; when’ (supposition, hypothetical, conditional)

Gea in Nmbo is a particle of supposition. Depending on the context it can be variously described as a conditional or hypothetical ‘if’, or an indicative or temporal ‘when’. Athanasiadou and Dirven (1997) have proposed that the concept of supposition is a prototype concept that unites various subtypes of conditionals, and I use this label to capture the wide breadth of usage for this Nmbo particle.

(61) a.

ag	qrvar	srt	**gea**	k\amdo/nga	wao,	qrvar
coconut	leaf	leaf	when	3sg.A:pfv.fut/become	ripe	leaf
ä	k\wie/nga...					
FUT	3sg:pfv.fut/fall					

When the leaf becomes dry, the qrvar will fall.

WSEK1-, 00:06:48.740 - 00:06:53.209

b. [Maiwa is giving advice to children about how they should work hard and not be lazy.]

yao	**gea**	k\ägävu/tawnge	mlml-am	yao
NEG	**SUPP**	2nsg.A:pfv.phab/work	sweat-ERG	NEG

k-n\räme/tao	...	bm	ädi	dena=wt	
3sg.A>2sg.U:ipfv.rmpst:ven/make		2ABS	EMPH	DEM.mn=PURP	
n/m,	nne=vnar=wt.				
2sg.U:nphd/be	food=PRIV=PURP				

If you do not work, if you did not sweat [lit. sweat does not make you]... you will surely be without food.

WSEK1-G20170707-03MY02MQhm,00:08:27.753 - 00:08:32.466

4.3.2.5 Yta (frustrative)

Yta signifies that an action was attempted, but was ultimately unsuccessful in its realisation. It is very similar to what has been described as a *frustrative* in the Amazonian language of Aguaruna (Overall 2007:384), and in some Papuan languages (e.g. Manambu (Sepik, Manambu; PNG) Aikhenvald 2008:293, Marind (Anim, Marindic; Indonesia) Olsson 2017:407)

A clause containing *yta* will often immediately be followed by another constituent indicating that the desired action did not occur, e.g. with the use of a negative particle (example 62a), or a counterfactual clause (b). This is not, however, always the case (c).

(62) a.

bs	**yta**	ge	t\mbrto/tam,	yawa!	Kitara	
fire.ABS	**FRUS**	FOC	1pl.A>3sg.U:ipfv.ypst/stop	NEG	platform	

mngo,	mñutyu	se	ta\w-akrn/!
house.ABS	all	INTS	3sg.A>3sg.U:pfv.proxin/stop

We **tried** to stop the fire, but not enough! All of the houses were completely burned.

WSEK1-G20151015-06BurnedMN, 00:01:16.040 - 00:01:20.640

b. **yta** yna wén banban=an k\äwere/tawn, wndede
FRUS this tree shadow=LOC 1sg.A:ipfv.rmpst/grip CNTF
bä nayu t/maro.
3ABS far 3sg:rmpst/be

I tried to grip onto the shadow of the trees, but they grew far away.

WSEK1-B20150804-01SnakeBiteTY, 00:08:23.941 - 00:08:29.426

c. **yta** y\uvey/aem aragav=ama arng-h=t
FRUS 1du.A>3sg.u:pfv.pret/struggle river=PERL paddle-INF=ALL

We tried struggling to paddle down the river.

WSEK1-B20150804-01SnakeBiteTY 00:02:22.975 - 00:02:26.820

4.3.3 Other

4.3.3.1 Yao (negation)

Nmbo has one negative particle *yao*. It is used for negating both nominals and predicates, so can be characterised as a general negator of constituents. *Yao* can also be used as a negative response like English "no". The negative particle tends to occurs at the left edge of the phrase (example 63a for NP, (d) for V), and semantically modifies its head. The negative particle can, however, occur in other positions (63b and (c)).

There is an alternative form, *yawa*, of the negative particle (see also example 62a in the preceding frustratives section). This appears to be derived by the phonetic merger of negative *yao* with a following *ä*, a reduced form of the future particle *bä*. I have not investigated whether *yawa* has any distributional or semantic differences compared to *yao*.

(63) [Constituent negation]

[YS talks about how she moved out of her home to her husband's village. The [I was escorted] in the translation is mentioned in the preceding clause.]

a. **yao** yna hamba=t, kt=ano Tubägut.
NEG DEM.proxvillage=ALL DEM.prox=LOC Tubägu=ALL
[I was escorted to] Not this village, but to Tubägu over there.

WSEK1-B20170721-01Yayam09Finalh5, 00:07:46.870 - 00:07:49.203

b. hrare **yao** site ym, qév-qév mi
moon **NEG** light(ADJ) 3sg:nphd/be dark-dark CONT
ym
3sg:nphd/be
The moon is not bright, it's still dark.

WSEK1-B20170610-03JY03Casshm, 00:02:32.012 - 00:02:33.692

c. yndon **yao** bä k-n\a-ramo/n be
1sg.ERG **NEG** FUT 1sg.A>2sg.U:pfv.proxin:ven/APP-give 2sg.DAT
I will not give it to you.

WSEK1-G20170630-01GW03Casshm, 00:03:59.341 - 00:04:01.227

d. nne **yao** t\ne/tawn.
food **NEG** 1sg.A>3sg.U:ipfv.rmpst/eat
I did not eat food.

WSEK1-B20150804-01SnakeBiteTY, 00:13:22.570 - 00:13:23.873

4.3.3.2 Ge, Gs (Adverbial Demonstratives)

The particles *ge* and *gs* are adverbial demonstrative particles that provide local adverbial information. The form *ge* appears to be used for proximate locations, while *gs* is used for distal ones.

Ge and *gs* can both occur by themselves in independent intonation units. When they do, *ge* and *gs* have variant forms. The *ge* form is often realised as *gihe*, with a nasalised final vowel and accompanying pointing gesture. *Gs* has a variant

Figure 4.16: Gestural examples of *gihe* and *gsou*.

form *gsou*, which seems to be used when referring to distant entities, objects, or locations. *Gsou* is also often accompanied with a pointing gesture. Figure 4.16 shows uses of the variant forms with the accompanying gesture.

Ge and *gs* often occur before and after verbal predicates. They can also occur before or after the nominal demonstratives, *yna*, *kt*, and *dena* (examples 65). In both cases the particles function to draw the hearer's attention to the entity or predicate they precede. *Ge* can also function as a general subordinator (section

4.9.7). When functioning as a general subordinator, *ge* and *gs* also appear to function as topic markers, or *thematising* markers, as found in many Papuan languages (Reesink 1994, Heeschen 1998, De Vries 1995, 2006). The Chapter 8 is dedicated to discussing this phenomenon in Nmbo.

Preverbal position:

(64) a. yndo yna ag **ge** y\iye/yn.
1sg.ERG DEM.prox coconut.ABS DEM 1sg.A>3sg.U:pfv.pret/plant

This coconut **here** I planted it.

WSEK1-G20170721-03BZ00Coconuthm, 00:01:01.046 - 00:01:03.685

b. bä **gs** y/m Namna zi-e ge
3ABS DEM 3sg.U:nphd/be namna language-DAT DEM
n\oav/tat.
3pl.A:ipfv.nphd/speak

Those people there, they speak with the Namna language. [lit. They, **there** they are, with the Namna language they really speak.]

WSEK1-B20150813-03PerceptualHZ, 00:03:14.953 - 00:03:17.644

Following nominal demonstratives:

(65) a. w\uvu/tamng yna **ge**.
3pl.A>1sg.U:ipfv.prim/intend DEM.prox DEM

He meant to hit me.

WSEK1-B20150909-MissingMoneyMZ, 00:10:24.103 - 00:10:25.376

b. kona-we q\rmdä/nawt dena **gs**,
rope-DAT3pl.A>1sg.U:ipfv.rmpst/tie like.this DEM.mn DEM
yna=no=tio **ge**.
DEM=LOC=TEMPG DEM

They tied me like this with the rope, over here,

WSEK1-B20150909-MissingMoneyMZ, 00:14:39.309 - 00:14:41.270

4.4 Nominal Morphology

The bulk of this section will present the grammatical and semantic case markers in Nmbo. Before doing so we will cover the nonfinite marker *-h* (section 4.4.1), agent nominalisation (section 4.4.2), and reduplication (section 4.4.3).

4.4.1 Infinitive Marker *-h*

Uninflected verbs in Nmbo must occur in an infinitive form. The infinitive form of the verb takes the suffix *-h*, which then allows the addition of case markers such as the dative *-e*, purposive *=wt*, and originative *=mn* to name a few. These case-marked infinitive verbs occur as part of subordinate clauses; as a complement of a complement taking verb (example 66a), and an adjunct (b). Infinitives can also occur as semantic auxiliaries in *phasal constructions* (section 4.9.1).

The semantics of the case marked nonfinite verb is somewhat transparent as instances of metaphorical extension of the prototypical semantics provided by the case, e.g. *freng-h=t* (fix-INF=ALL) > do fixing on an object (example 66a).

(66) a.

ynd	bva	n\awatambne/tan	inzin	**freng-h=t**.
1ABS	just	1sg:ipfv.nphd/learn	engine.ABS	**fix-INF=ALL**

I just learned how to fix engines.

Fieldnotes 2017:9

b.

bä	t/utawn,	ärgv=at	**on-h=t**...
3ABS	3sg.U:rmpst/walk	creek=ALL	**fishing-INF=ALL**

She walked to the creek to go fishing.

WSEK1- B20150805-01NinyiTS, 00:01:01.355 - 00:01:02.300

4.4.2 Agent Nominaliser

Verbs can be agent nominalised by the suffix *-här*. This suffix likely arose from a historical compound of the word *är* 'man; person' following a non-finite verb i.e. *waovaro-h*, 'to teach someone' > *waovaro-här* 'teacher'; *lmä-h* 'to pull something' > *lmäne-här* 'leader'. It appears to be a semi-productive suffix, with

some words being creatively derived in the moment (*olto-här* in example 67a), while other words are thought to be lexicalised. Forms with non-compositional semantics given the root verb (e.g. *ägävu-här* in 67a) are also thought to be lexicalised.

Agentive nominalisations are often found in relative clause constructions (section 4.9.7), where the word is followed by the relativiser *är* 'man; person' (contrast examples 67b and c). This suggests that *-här* is no longer compositional, and further suggests that some agentive nominalisations may be treated as phonetically and morphologically compressed (i.e. on a spectrum of lexification).

(67) a. [Bunai is telling a story about the time he was on board a dingy that sank. In the following utterance he is characterising himself and his travel companion.]

dena	ägävu**-här**	är	yn/m,	ynd.
DEM.mn	work**-AG.NMZR**	person.ABS	1pl.nphd/be	1ABS
Olto-här	är.			
sink-AG.NMZR	person.ABS			

We are such kind of people. People who have sunk.

WSEK1-A20150917-1SinkingBA, 00:22:58.870 -00:23:00.920

b. [Maiwa is talking about the day he got married, and how his age-mates were responsible for brining the food.]

tande	nne	rsa**-här**	ta/maro	Bévék	[...]
1sg.POSS	food	bring**-AG.NMZR**	3pl:rmpst/be	Bévék	[...]
Nuwira,	Masa				
Nuwira,	Masa				

My food bearers were Bévék, Nuwira, and Masa.

WSEK1-G20170713-02MQMW, 00:09:00.083 - 00:09:01.580

c. [Smako is talking about how he and others brought Christianity to the Kerake area.]

ynd	misin	rsa**-här**	är	bä	yn/maro.
1ABS	mission	bring**-AG.NMZR**	person	3ABS	1pl:rmpst/be

We were the bringers of Christianity [lit. we were the people who were bringers of the mission]

WSEK1-B20150930-01ChurchSZ, 00:09:53.063 - 00:09:54.979

4.4.3 Reduplication

Nmbo nominals can be reduplicated, both fully and partially. The most productive process of reduplication is full reduplication where a word form is copied completely. Partial reduplication is a peripheral process often found with bird noise mimicry (e.g. *serere* for the call of the *kazn serere* 'Emperor fairy wren', the *wambo sansa* 'Willy Ragtail' who cries 'san-san'). This section will cover the functions of full-reduplication.

The reduplication of nominals has multiple functions. Reduplication of common nouns tends to create a plural meaning, e.g. 'many X' (examples 71a and b), occasionally 'a small version of X' (*sq* 'canoe', *sq-sq* 'toy canoe'), or sometimes both (*séng-séng*, 'lots of small things' such as insects, birds, children). Reduplication is not obligatory for plural referents.

(68) a.

mrz-mrz-am	w\mdo/st		ynd	dena	de...
girl-RDP-ERG	3nsg.A>1sg.U:pfv.pst/tell		1ABS	DEM.mn	ALR
"Kaki,	helokovta	bä	n\avn/d."		
grandparent	helicopter.ABS	3ABS	3sg.A:ipfv.nphd/fly		

The girls told me like this ... "Granny, the helicopter, it is flying."

WSEK1-B20150813-05PlaneMZ, 00:01:02.396 ... 00:01:12.886

b. yäme **hara-hara**, bés hara, bés gane=va, hi
mat **bundle-RDP** fire bundle fire bark=COM torch
sv-sv, ynd yna=va.
bark-RDP 1ABS this=COM

Bundles of mats, a fire starter bark bundle, torch barks... I was there with these things.

WSEK1-B20150813-05PlaneMZ, 00:01:04.850 - 00:01:10.500

Another function of reduplication is the naming of flora, fauna, and other natural phenomena. The reduplication of one type can produce the name of another type, e.g. reduplicating the name of a plant produces the name of some animal (table 4.18). This process of reduplication is not a productive pattern per se, since there are examples of inherently reduplicated names with no corresponding non-reduplicated form, e.g. *dar-dar* (a general 'butterfly') but **dar*.

Base	Reduplicated
Kakma: Horn-billed bird of the classification *Rhyticceros plicatus.*	*Kakma-kakma:* A plant of the classification *Cupaniopsis.*
Karv: Anthill	*Karv-karv:* A general word for 'goanna'.
Kave: Silver crested cockatoo.	*Kave-kave:* Crouton plant.
Sov: Wave	*Sov-sov:* A kind of tree that grows in the bush.
Wagiv: A general word for fish.	*Wagiv-wagiv:* Tree of the classification *Litsea firma.*

Table 4.18: A selective list of flora and fauna names with reduplicated and corresponding non-reduplicated forms.

The reduplication of adjectives produces an effect of intensification. It is not a comparative construction, since the reduplication is not inducing a direct comparison of states e.g. *mngar* 'quick', *mngar-mngar* 'go faster!' ; *kitong* 'large', *kitong-kitong* 'very large'.

Nominals other than common nouns can also be reduplicated, though some of these are likely fully lexicalised rather than part-taking in a synchronically productive process. For example the word *ka-ka* meaning 'near' looks like a reduplicated form of the indefinite pronoun *ka* 'where', but I have not found other redu-

plications of indefinite pronouns. The quantifier *ämb* 'some' is also reduplicated to produce *ämb-ämb-e* meaning 'for good' or 'forever' (section 4.2.1.5).

(69) [Taqm is telling a story about the men's initiation ritual. In one part of the ritual the men would frighten women by making loud noises outside the houses where the women hid.]

a. bä ama=va fyo ä ta/m, bä **ämb-ämb-e**
3ABS mother=COM then FUT 3nsg:ypst/be 3ABS **some-RDP-DAT**
är mngo=t ädi bva n\oga/nat.
person.ABS house=ALL EMPH just 3nsg.A:ipfv.nphd/clear.out

Having been done with the mothers, they [the men] completely clear out from the house.

WSEK1-B20170628-Ari4712, 00:01:47.660 - 00:01:50.600

(70) [Kawas is telling a story about the time an elderly man passed away.]

a. foyo **ämb-ämb-e** band=an ge n\akmo/i.
then **some-RDP-DAT** ground=LOC DEM 3sg.A:pfv.pret/lay

Then, he was lying on the ground for good/forever.

WSEK1-B20150813-02DimbanKSae, 00:09:18.971 - 00:09:20.725

Finally, the manner demonstrative *dena* 'like this' (section 4.2.1.3) can be reduplicated. The usage of this form is akin to English 'etc etc'. The use of *dena-dena* indicates continuation while also indicating the content of the continuation is not particularly important.

(71) a. [Thomas came back from Balimo yesterday, and is telling about who he spoke to yesterday.[8]]

[8]This is an interesting construction where the predicate of the first clause has a dual thematic *-w* indicating the recipient argument is dual number (Eno and another person, indicated by the plural dative.)

Eno-ve=t t-ng\wa-waito/w-n yndon [...]

Eno-DAT.nsg=ALL 1sg.A>3pl+.U:ipfv.ypst/APP-tell 1nsg.ERG ...

skul=mn zi-zi, dena de skul kt

school=ORI story-RDP DEM.mn ALR school DEM.dist

n\ovro/wm ynd **dena-dena**....

1du.A:ipfv.nphd/do 1ABD DEM.mn-RDP

I told many stories to Eno [and some other person who was present]... Stories from school, the two of us do like this at school, etc etc...

WSEK1-B20170627-01YpstTThm, 00:03:47.000 - 00:03:52.906

4.4.4 Case Markers

Nmbo has grammatical case markers and a rich set of semantic case markers. A summary of the terms, forms, and functions, are presented in table 4.19. The following descriptions of case inevitably makes reference to the verbal complex of Nmbo; terminology referring to the parts of the verb complex are presented in section 4.5.

Following the definition of Blake (1994), Nmbo case is defined as encoding core arguments subcategorised for and indexed by the verb (p.1). Since Nmbo verbs can index the existence of a third argument by the applicative prefix (section 4.5.2), this third argument constitutes a core argument of Nmbo. The core arguments are represented by the letters S, A, P, and R: S is the sole argument subcategorised by a one-place verbal predicate; A is the agent-like argument subcategorised by a two-place predicate; P is the patient-like argument subcategorised by a two-place predicate; R, the recipient, is the third argument of a three-place predicate (c.f. Comrie 1978, Haspelmath 2011). The grammatical case markers in Nmbo are therefore the following: *ergative* (ERG), *absolutive* (ABS), *dative* (DAT), and *possessive* (POSS).

The precise scope of the case markers have not been investigated, and there appear to be fine-grained variations of behaviour and semantics depending on the subclass of the head nominal. Subsequently it is still unclear how to best analyse these bound morphemes; i.e. as affixes or clitics. For the purposes of this sketch grammar, the grammatical case markers are all treated as suffixes that occur at

Label	Gloss	Semantic Role	Grammatical Relation	Verbal Index	Form
Grammatical					
Absolutive	ABS	Sole argument of monovalent verb.	S	A/U U	ø
		Patient of transitive verb,	P	U	
Ergative	ERG	Agent of divalent verb.	A	A	**-m /-vem***
Dative	DAT	Recipient of transfer verb, instrument, source location	R	(U)**	**-e / -ve***
Possessive	POSS	Possessor	Modifier	(U)	**-ende/ -vende***
Semantic					
Goal	GOAL	Goal of transfer verb.			**=vav/ =vevav***
Benefactive/ Purposive	BEN/PURP	Purpose, intention, of movement			**=wt**
Locative	LOC	Location		No	**=n / =(a)no****
Allative	ALL	Locational goal, goal of motion		No	**=t**
Perlative	PERL	Motion through		No	**=ama**
Dem. Ablative	DEM.ABL	Source location		No	**=anma**
					=(a)n(d)ma
Temporal	TEMP	Temporal location,		No	**=tawa,**
		'During X time'			**=taw**
Temporal Goal	TEMPG	Temporal goal,		No	**=tio**
		'Until X time/moment'			
Comitative	COM	Accompaniment		No	**=va**
Privative	PRIV	Absence		No	**=vna(r)**
Similative	SIM	Comparison		No	**=nit**
Originative	ORI	'Source, reason'		No	**=mn**
Restrictive	RST	'Only X'		No	**=ro**
Alone	ALONE	'X alone'		No	**=yo**

Table 4.19: Nmbo case marker overview. Alternative forms will be presented in the relevant sections. * indicated the non-singular form when attached to a human referent, or the dative form of the pronoun; ** = forms demonstrative paradigm ; (U) indicates variability, with indexing on the undergoer suffix being optional

the right edge of an NP (contrast examples 72a and b). A morpheme having scope over a clause or phrase is a criterion often raised for distinguishing clitics from suffixes (e.g. Zwicky and Pullum 1983, Aikhenvald 2002b), but grammatical case markers in particular have been recognised as having scope over a wider domain than just the head (c.f. Dench and Evans 1988:3-6). Nmbo grammatical case markers are the left-most bound inflectional morpheme, and can be followed by other bound morphemes, such as the restrictive *=ro* (example 72c, the form of the restrictive is *=ru*). This suggests that grammatical case markers are in some sense more bound to their argument in a way the semantic case markers are not.

(72) a. [kruvr**-am**]$_{NP}$ fivi ädi w\rame/t.

cold-ERG really INTS 3sg.A>1sg.U:ipfv.nphd/give

The cold is really really getting to me. [lit. The cold is really really giving me.]

WSEK1-G20170723-01GwN10KK, 00:03:05.771 - 00:03:07.310

b. [kruvr gänzän**-am**]$_{NP}$ w\vn/d.

cold gigantic-ERG 3sg.A>1sg.U:ipfv.nphd/bring

The gigantic cold is making me sick. [lit. The gigantic cold is bringing me.]

WSEK1-G20170723-01GwN10KK, 00:03:27.010 - 00:03:30.778

c. är**-m**=ru t-ng\ne/tawt.

people**-ERG**=RST 3pl.A>3pl+.U:ipfv.rmpst/eat

Only the old ladies ate these. [Note: Contextually *är* is understood as old ladies.]

WSEK1-B20150727-01PastCustomMZ, 00:00:45.400 - 00:00:46.600

Semantic case markers in Nmbo indicate the semantic role of a participant which is not a core grammatical role. Constituents that are marked with semantic case markers are therefore often adjuncts or complements in a subordinate clause. All semantic case markers are analysed as enclitics, since they can occur to the right of the grammatical case markers (examples 73a, b).

(73) a. [Richard is answering the question “Do you catch crocodiles?”]

yao,	ana-vem=ro	y-ng\ere/tat.	yndo
NEG	elder.sibling-ERG.nsg=RST	3pl.A>3pl+:ipfv.nphd/catch	1sg.ERG
kor	yao.		
really	NEG		

No, only my elder brothers catch them. I don’t really.

WSEK1-G20150825-01CoconutRGMN, 00:06:54.948 - 00:06:57.241

b. [Bézbéz is explaining how her husband sent a child with a message to Embiak.]

foyo	ämb	toge	y-n\a-vare/tam,	Embiak-e=vav
then	some	child	3sg.A>3sg.U:ipfv.prim:ven/APP-send	Embiak-DAT=GOAL

Then he sent some child to Embiak.

WSEK1-B20150721$_0$1marriageAT, 00:11:30.732 - 00:11:33.320

Some case markers exhibit phonological and lexical allomorphy. An example of phonologically conditioned allomorphy is the locative *=an* gaining an epenthetic palatal to break VV sequence, e.g. *tomba* ‘long’ + LOC producing *tombayan*. An example of lexical allomorphy includes the variant forms of the dative *-e*, such as *toge* (‘child’) + DAT producing *togai*, and *dmave* (‘wife; woman’) + DAT producing *dmaveai*. Known allomorphic forms will be presented in each of the relevant sections.

Personal pronouns form paradigms with most of these case markers. For morphemes unique to personal pronouns (i.e. the close possessive *-nzo* and ‘alone’ suffix *-mbia*) see section 4.2.1.1.

4.4.4.1 Ergative -m, and Absolutive ø

The ergative case marks the semantic role of actor in a two-place predicate. It is expressed by the suffix *-(á)m*, *-am*, or occasionally *-um* or *-om*. The non-singular form is variously *-vam* or *-vem*, and can only be taken by human nouns. The

person and number of the ergative-marked noun is indexed on ambifixing verbs by the a-suffix.

(74) a. Doado-**m** bavwa y\ne/**t**.
Doado-ERG taro.ABS 1sg.A>3sg.U:ipfv.nphd/cook
Doado is cooking a taro.
Field notes Book 1, 2017:12

b. ama-**vem** mamwi ge y\r/**st**.
mother-**ERG.nsg** pig.ABS DEM **3nsg.A**>3sg.U/carry.pfv.psta
The mothers carried the pig.
(Karat, 00:02:18.664 - 00:02:19.991)

The absolutive is the unmarked form (ø). It encodes the sole argument in one-place predicates (example 75a), and the P argument of two-place predicates (75b).

(75) a. tande **ama** **t**-maren/grwn, ämbru wik.
1sg.POSS **mother.ABS** 3sg.U:rmpst/stay one week
My mother stayed there, one week.
WSEK1-G20170617-01Alqi02MQ, 00:01:09.443 - 00:01:12.780

b. tande **ama** **t**\ake/tawn.
1sg.POSS **mother.ABS** 1sg.A>**3sg.U**:ipfv.rmpst/watch
I was watching my mother.
(WSEK1-G20170722-02GwN00Cocohm, 00:02:18.664 - 00:02:19.991)

Human and common nouns take the ergative, but abstractions such as emotions or sensations like *bérbér* 'fear', *dmke* 'pain', and *kruvr* 'the cold' may also take the ergative and function as the agent of an *experiencer object construction* (section 4.8.2.1). The person experiencing the sensation is in a patient role in the absolutive, and indexed on the verb by the u-prefix.

(76) a. ynd bérbér**-am** q\rame/**tao**

1ABS fear**-ERG** **3sg.A**>1sg.U:ipfv.rmpst/give

I was scared. [lit. Fear came to me].

(WSEK1-B20150813-02HomHZ, 00:01:41 - 00:01:42)

b. kruvr**-am** de w\ivo/ø

cold**-ERG** ALR **3sg.A**>1sg.U:pfv.pst/finish

I was cold. [lit. The cold finished me off.]

(WSEK1-G20170723-01GwN10KK, 00:02:30.412 - 00:02:31.247)

4.4.4.2 Dative -e

The suffix *-e* (or *-ye*, *-ae*) fulfils numerous functions, which are analysed as multiple functions of a single dative suffix rather than positing multiple homophonous suffixes.

The dative *-e* can mark the recipient argument R of a transfer verb. This referent is then indexed on the verb in one of two ways. One way is to indicate the addition of an R argument by the benefactive/applicative prefix *(w)a-* immediately left of the verb root. The benefactive/applicative prefix does not express person and number, but merely indicates the presence of a third argument (example 77a). For more details on three-place predicates, see section 4.8.3.

Sometimes, however, the dative-marked R is indexed on the u-prefix, showing agreement of person/number. The theme argument T (i.e. the object being transferred) is then indicated on the verb by the applicative (example 77b). Note that in example 77a dual number theme *sombwi nne* 'two yams' is not indexed on the verb in any way. In example 77b the applicative prefix flags the presence of T, while the diathetic right of the verb root is in the dual form *-we*.

(77) a.

yndo	Eri**-e**	sombwi	wagiv	y-**a-**ram/we-n.
1sg.ERG	name**-DAT**	two	fish	1sg.A>3sg.U+DU:ipfv.nphd/**APP**-give
A	R		T	

I gave Eri two fish.

Fieldnotes 2014 *The Book of Collections* :26)

b. ymo sombwi nne w\a-ram/ø.
3sg.ERG two yam.ABS 3sg.A>1sg.U:pfv.pst/APP-give
A T
He gave me two yams.

Fieldnotes 2014 *The Book of Collections*: 7)

The dative can also mark the semantic role of *instrument*. The instrument is usually not indexed on the verb by the applicative, but it is seemingly possible to do so (example 78b).

(78) a. Bä ta-ng/maron, kt=ano Wando, baskol**-e**.
3ABS 3pl:rmpst/go DEM.dist=LOC Wando bicycle**-DAT**
They went to Wando, using a bicycle.

WSEK1-B20150721-01marriageAT, 00:11:42 - 00:11:47

b. Yndo kai t**a**-vo/tan wén busäl**-e**.
1sg.ERG ± 1day 1sg.A>3sg.U:ipfv.ypst/**APP**-fell tree.ABS axe**-DAT**
Yesterday I felled a tree with an axe.

Fieldnotes 2014 *The Book of Collections*: C

The dative sometimes functions with *ablative* semantics, i.e. indicates the source location of movement. This interpretation seems to be limited to referents that are locational proper names. Nmbo does not have a dedicated ablative morpheme, unlike her sister languages Nen and Nama (*-ngama* and *-mé* respectively). The dative marked referents used in an ablative sense are not indexed on the verb.

(79) a. Balimo**-e** woi n-n\tor/ym.
Balimo**-DAT** again 1pl:pfv.pret:ven/exit
From Balimo we came out again.

WSEK1-B20150818-03BalimoKS, 00:07:53.577 -00:07:57.090

Finally, dative *-e* suffixed on infinitive verbs produce manner adverbial semantics. Note that the same function in Nen is carried out by a dedicated suffix *-ae*, which is different in form to the Nen dative suffix *-gta* or *-vav*. Given this, it is possible that the manner adverbial semantics of Nmbo *-e* historically descended from another suffix, and is synchronically homophonous with the dative *-e*. Nonetheless the dative *-e* is analysed as providing manner adverbial-like meaning since there is a similarity of semantics between instruments and manner adverbials (example 80).

(80) a. emo y-ng\ane/i oro-h-ye.
who.ERG 3sg.A>3sg.U:pfv.pret:and/take steal-INF-DAT
Who stole it? [lit. Who took it with stealth/stealingly.]

WSEK1-B20150928-MPVisitLS, 00:03:17.781 - 00:03:19.153

b. ymo tande naifa mrsa-h-e
3sg.ERG 1sg.POSS knife.ABS take.without.permission-INF-DAT
y-ng\ane/.
3sg.A>3sg.U:pfv.pst:and/take
He took my knife without permission. [lit. He took it without authority/unauthorisedly.]

Nmbo Dictionary Draft Version 2018, entry *mrsah*

4.4.4.3 Possessive -ende

The possessive suffixes *-(e)nde* (singular) and *-vende* (non-singular) mark the semantic role of possessor. Human nouns and common nouns can both host the possessive suffix, though only the former can host the non-singular.

(81) a. ama**-nde** ytqn eve y/m?
mother**-POSS** name who.ABS 3sg:nphd/be
What's mother's name? [lit. who is mother's name?]

Fieldnotes 2015 Book 1: 22

b. ama-**vende** knd=n mé q/maron

mother-**POSS.nsg** inside=LOC CONT 1sg.U:rmpst/be

I was still at my parents' place.

(WSEK1-G20150902-07MarriageSY2, 00:06:25 - 00:06:27)

The person/number of the possessive pronoun can be indexed directly on the verb. In the following example, *tande hkv* 'my eyes' with the first singular possessive pronoun, is agreeing with the u-prefix *w-* on the main verb.

(82) [Bunai is explaining how salt water was getting into his eyes as his dinghy began sinking.]

a. [**tande** hkv mñutyu] [rar-h ynao **w**-n\av/ngo]

1sg.POSS eye.ABS all bite-INF here 3sg.A>**1sg.U**:pfv.fut:ven/begin

My eyes were beginning to hurt all over. [lit. My eyes all were beginning to be bitten (by the salt)]

WSEK1-A20150917-1SinkingBA, 00:12:20.770 - 00:12:23.917

4.4.4.4 Goal =vav

The goal marks a human destination of a motion. The goal marker follows a dative marker *-e*, and thus the argument may be optionally indexed on the verb by the benefactive/applicative. Non-human destinations are often marked by the allative *=t* (section 4.4.4.7).

(83) a. yndo buk y-ng\rsa/n waovaro-här-ue=**vav**

1sg.ERG book.ABS 1sg.A>3sg.U:pfv.pst:and/carry teach-AG.NMZR-DAT=**GOAL**

I carried the book to the teacher.

WSEK1-B20170627-03SmakoSL02Finalhm, 00:01:27.788 - 00:01:30.611

b. k\ango/tawn mngo, mwidadm-ae=**vav**

1sg.A:ipfv.rmpst/return house.ABS exchange.aunt-DAT=**GOAL**

I returned home, to my auntie.

WSEK1-B20170627-03SmakoSL02Finalhm, 00:01:27.788 - 00:01:30.611

The goal marker likely arose from the lexical noun *fav* meaning 'place'. Indeed compounding *fav* with another lexical noun often productively produces names for locations with specific function i.e. *anu-vav* 'washing place', *kmh-vav* 'sleeping place'. One way to determine whether the segment *vav* is compounding or case marking is the presence of the dative *-e* preceding *vav*. Example 4.4.4.4 shows two cases of compounding; (a) with an instance where *vav* is not preceded by the dative, and (b) with an infinitive verb compounded without the presence of the dative.

(84) a. trak t-ng\rs/ao kt=ano... mora-mora

truck 3sg.A>3sg.U:ipfv.rmpst:and/run DEM.dist=LOC medicine-RDP

mngo**-vav**...

house**-place**

(He) ran the truck all the way there... to the hospital...

WSEK1-G20150902-03PigGS, 00:19:15.170 - 00:19:18.550

b. yvende wumeng-h**-vav** ka y/m?

3nsg.POSS meet-INF**-place** where 3sg.be/nphd

Where is their meeting place?

WSEK1-G20150826-07CocoIntZG, 00:09:13.534 - 00:09:15.863

4.4.4.5 Benefactive/Purposive =wt

Nmbo *=wt* has a range of semantic functions depending on the noun it binds to. When the host is a human noun, the benefactive/purposive expresses the

semantic role of beneficiary. Like the goal marker (section 4.4.4.4), *=wt* attaches to the dative form of the pronoun (example 85a), or follows the dative suffix *-e* (85b). Subsequently the marked nominal is optionally indexed on the verb by the applicative. In example 85a the main verb has the applicative prefix indexing the theme argument (the dinghy), while the 2nd singular R argument is indexed on the u-prefix. The applicative is absent in example 85b.

(85) [Benefactive use.]

a. Yndo be=**wt** ä k-n**a**-ramo/n...
1sg.ERG 2sg.DAT=**BEN** FUT 1sg.A>2sg.U:pfv.proxin:ven/**APP**-give
nu=ama wrse-här rokar.
water=PERL do-AG.NMZR object
I will give you... a dinghy [lit. an object that goes across the water].

WSEK1-B20150928-MPVisitLS, 00:04:22.755 - 00:04:26.360

b. wrar ge y-n\nam/da, bä
animal.ABS DEM 2sg.A>3sg.U:ipfv.nphd/shoot FUT
y-n\rs/ deva-ve=**wt**.
2sg.A>3sg.U:ipfv.nphd:ven/deliver father-DAT.nsg=BEN
If you shoot an animal, you would present it to the fathers [You will give them the meat for their benefit.]

Fieldnotes 2017: 112

A common noun with *=wt* signals the the purpose of the predicate verb. The benefactive/purposive can be used to signal intention, such as the intent to clearing the garden for the purpose of planting yams (example 86a), or going somewhere for the purpose of getting medicine (86b). The dative is seemingly absent when *=wt* is used as a purposive marker, indicating the adnominal characteristic of benefactive/purposive marked common nouns.

(86) [Purposive use.]

a. yu t\rte/tawm **nne=wt**

space 1pl>3sg.U:ipfv.rmpst/clear food=**PURP**

We cleared the space, for food [e.g. planting food]

WSEK1-B20150813-01RoyJY, 00:00:14 - 00:00:16

b. kt y-n/maro mora-mora=**wt** yävv

DEM.dist 3sg:prim/come medicine-RDP=**PURP** 3nsg.DAT

I came there to them for medicine.

WSEK1-B20150909-MissingMoneyMS, 00:04:38 - 00:04:41

Infinitive verbs with a following *=t* provide a 'going to do X activity' interpretation. In example 87a the spears are made for the purpose of shooting the ogre (lit. 'going to make a spear'). In example 87b the mark was set for the purpose of scheduling a fishing expedition (lit. 'going to fishing'). I have not come across the form *=wt* when the host is an infinitive verb, making this homophonous with the allative *=t* (section 4.4.4.7). Since there is much semantic overlap between 'going for the purpose of doing X' and 'going towards X', it is perhaps more accurate to describe the benefactive/purposive as syncretic with the allative when the host is an infinitive verb.

(87) a. trmr hara t-ng\avro/tawt, yna

spear bundle 3pl.A>3pl+.U:pfv.rmpst:/make DEM.proxogre-DAT

niña-e wävätu-h=**t**

shoot-INF=**PURP**

They made spears bundles to shoot that ogre.

WSEK1-B20150805-01NinyiTS , 00:11:55 - 00:11:57

b. fraidei mak y\uvu/tamm, on-h=**t**.

Friday mark 1pl.A>3sg:ipfv.prim/set fishing-INF=**PURP**

We set the date as Friday to go fishing.

WSEK1-G20150921-LostZS, 00:01:09 -00:01:12

Demonstratives can take the benefactive/purposive, but when they do, the demonstratives must first be marked with the locative *-(a)no*. The benefactive/purposive will follow to the right, e.g. *yna=no=wt* here=LOC=BEN/PURP, **yna=wt*. When the demonstrative is marked with the benefactive/purposive, it is used in a pronominal sense, i.e. these people over here (example 88).

(88) a. Yna nne, kt=ano=wt y/m, Bevdvn.
DEM.prox food DEM.dist=LOC=BEN 3sg.U:nphd/be Bevdvn
These yams, it's for there, for Bevdvn.

Fieldnotes 2017: 112

4.4.4.6 Locative =n

The locative *=(á)n*, *=an*, or *=yan*, provides general locational information such as being at, or on, a place or thing. The locative suffix can occur on common noun(89a), proper nouns of location (e.g. Port Moresby, example 89a), and temporal nouns, e.g. *qévte* 'night' (90). The locative, like the core case markers, occur at the right-edge of the NP (89c). Locatives are also found on spatial demonstratives, and have a slightly different form *-(a)no* (91). The proximate demonstrative occasionally takes the form *ynan* in the corpus, and is treated as equivalent to the more common *ynano*.

(89) a. ynd mosvi=**n** q/maro
1ABS Moresby=**LOC** 1sg:rmpst/be
I was at [Port] Moresby.

WSEK1-G20150831-05WorkingMG, 00:02:56.500 - 00:02:58

b. yndo säläme y\vträr/yn handir=**an**
1sg.ERG cloth 1sg.A>3sg.U:pfv.pret/take.off banksia=**LOC**
t\vyo/n
1sg.A>3sg.U:pfv.proxin/hang
I took the cloth off and put it on the banksia tree.

WSEK1-B20150924-HuntingWZ, 00:00:31-00:00:35.500

c. gea t\inge/aem, [gngn kitong=**an**] t/lao-ngrwn.
SUP 1du.A>3sg.U:pfv.pret/see bush.type big=**LOC** 3sg:rmpst/be.inside
When we found it, it was in the large bush.
WSEK1-G20150902-03PigGS, 00:07:04.800 - 00:07:07.380

(90) [Temporal use]

a. ḡḡn t\akavo/n qévt=**an** ädi n\owano/yn
bow 1sg.A>3sg.U:pfv.proxin/get night=**LOC** EMPH 1sg:pfv.pret/travel
I got my bow... and went out at night.
WSEK1-G20150902-03PigGS, 00:01:52.500-00:01:56

(91) [Spatial demonstrative use. Note the alternate use of *yna=no*.]

a. markai ge n\ovar/eng **yna=no**...
white.people DEM 3pl.A:pfv.pret/arrive DEM.prox=LOC
The white people arrived here a long time ago...
02ClothesFTS, 00:00:00.441 - 00:00:02.4416)

b. bä Tekam **ynao** t-ng/m fronde Savaram
3ABS Tekam **DEM.prox.LOC** 3sg.ypst/go first Savaram
e~, **kt=ano** Arovwe.
DISC **DEM.dist=LOC** Arovwe
Tekam, he went from here, first from Savaram all the way~ there to Arufi.
WSEK1-B20150813-05PlaneMZ, 00:02:35.656 - 00:02:39.913

4.4.4.7 Allative =t

The allative *=(á)t* signifies motion towards a destination. The hosts of the allative are mostly locational nouns, and the destination argument is not marked

on the verb by the applicative. This makes the behaviour of the allative distinct to that of the benefactive which can verbally mark the third argument. I have separated the benefactive/purposive and allative as distinct categories because the host nouns are seemingly different in kind, but as alluded to in the benefactive/purposive section, the linguistic distinction between the allative, benefactive, and purposive, are somewhat fuzzy.

The *=(á)t* prototypically occurs on referents that are non-human destinations, and signals motion towards that destination. Motion towards a human destination is marked by with the goal *=vav*, e.g. *Bunai-e=vav* 'towards Bunai' (section 4.4.4.4). In elicitation sessions I have found that pronouns can take the allative form after the goal *=vav*, e.g. *yä=vav=t* 'towards him', but this form could not be found in the current form of the Nmbo Sociolinguistic Corpus.

The form of this suffix is usually *=t* when following a word ending in a vowel, but alternate forms such as *=yet* has also been observed in the same environment (example 92b). Other forms include *=(y)at*, *=(y)et*, *=(y)ot*.

The allative can occur on the proximal demonstrative *yna* to form *yna=t*. Speakers rejected possible distal demonstrative forms, **kt=t*, **kt=yt*, **kt=wt*. The word *yna=t* means something like 'at that place' but with an accompanying motion. In example 93, a better translation capturing this sense might be "Luckily Kaog was there as I looked towards that location". When *yna=t* occurs clause initially, it seems to be used with a meaning akin to 'after that'. Clause-initial *yna=t* frames the following clause as occurring immediately after, or because of, the preceding clause (example 94).

(92) a.

n-ng\itr/em	kt-andmae	Baono-ende	mngo=t
1du:pfv.pst:and/move	DEM.dist-DEM.ABL	Baono-POSS	house=ALL

The two of us moved away from there to Baono's house'

WSEK1-B20150909-MissingMoneyMS, 00:04:45 - 00:04:49

b.

är	ge	y-ng/m	dumgu=yat...
man.ABS	DEM	3sg:nphd/be	hunting.hut=ALL...

A man goes to the hunting hut...

WSEK1-B20150820-cassowaryLS, 00:01:04 - 00:01:06

(93) [With spatial demonstrative]

a. [Ziniki is telling a story about the time she got lost in the jungle, and here is telling about the moment she found other people.]

kt=ano		n\ongovary/n	ew\ake/tayn,
DEM.dist=LOC		1sg.A:pfv.pret/arrive	1sg.A>3nsg.U:pfv.pret+ipfv/see
kanzo	Kaog	yi\ngo/n	**yna=t**.
lucky	Kaog	1sg.A>3sg.U:pfv.pst/find	**DEM.prox=ALL**

I arrived there and saw them, it was lucky I found Kaog there. [alt: Luckily Kaog was there as I looked towards that location.]

WSEK1-G20150921-LostZS, 00:02:45.616 - 00:02:47.179

(94) [With spatial demonstrative: 'and then' sense]

a. [G̃ño is telling a story about two birds who were playing pranks on each other. The birds had stolen each other's clothing, and are now making a truce.]

ädi	n\ovamne/aend.	**yna=t**	bä	yänzu
EMPH	3du.A:pfv.pret/exchange	**DEM=ALL**	3ABS	3sg.POSSC
säläme=va	kakayam	ym	totr	
clothes=COM	bird.of.paradise	3sg:nphd/be	now	

[Having stolen each other's clothes] they exchanged [the clothes] back. Now the Bird of Paradise has her own clothes back.

WSEK1-G20170723-01GwN10KK, 00:05:01.155 - 00:05:02.218

4.4.4.8 Perlative =ama

The morpheme *=ama* expresses motion through space. The motion may be through something such as a jungle (example 95a), or going across the surface

of something (b).

(95) a. kt n\uvartu/yn ge ta/mn orang=**ama**
DEM.dist 1sg.A:pfv.pret/veer DEM 3pl:ypst/be jungle=**PERL**
k\ilau/en
1sg:pfv.proxin/enter

There I went off road... then I entered into the forest... [lit. I entered through the forest]

WSEK1-G20150826-01CowWG, 00:01:47 - 00:01:51.600

b. nu kt ä t\anwane/nge mna=t,
liquid DEM.dist FUT 3pl.A>3sg.U:pfv.fut/take DEM.fw=ALL
nu=**ama** ge n\rsa/t, dingi-e
water=**PERL** DEM 3sg.A:ipfv.nphd/run dinghy-DAT

They will buy fuel there for travelling on the water [lit. through the water], for the dinghy.

WSEK1-A20150922-10TzRT, 00:02:49.600 - 00:02:51.500

4.4.4.9 Demonstrative Ablative

The demonstrative ablative *-ama* suffixes to both proximal and distal spatial demonstratives. The name is a slight misnomer, as it also occurs on the manner interrogative *dr* 'how did it happen; what happened' (section 4.2.1.6). The demonstrative ablative has a variety of forms (summarised table 4.20), and Bevdvn villagers report that the form with the stop /d/ (e.g. *yna-n**d**ma*) is unique to Govav speakers. A cursory look at the corpus, however, suggests that both forms are used by both villagers.

The various forms suggest a compositional past, where the suffix is comprised of a locative (*-n*), and perlative (*-ama*) with an optional dative at the end (*-e*). Nmbo does not have a dedicated ablative suffix, but Nen has an ablative *-ngama* which suffixes to common nouns as well as spatial demonstratives.

(96)

Proximal	Distal	Interrogative
ynama		
yna-nma	kt-anma	dr-anma
yna-ndma	kt-andma	dr-andma
yna-ndmae	kt-andmae	dr-andmae

Table 4.20: Ablative demonstrative forms which have been observed.

a. [Gerida is explaining what can be done with a coconut leaf.]

tivravr	ä	y\avars/an	**yna-ndma**.
broom	FUT	1sg.A>3sg.U:ipfv.nphd/design	DEM.prox-DEM.ABL

I will design a broom from this.

WSEK1-G20141008-02CoconutEN, 00:07:18.097 - 00:07:20.119

b. [Gerida answers a question about the origins of her mother's mother.]

yände	ama	**kt-andmae**	y-n/maro,	Dimsisi.
3sg.poss	mother	**DEM.dist-DEM.ABL**	3sg.A:prim/come	Dimsisi

Her mother comes from that place, Dimsisi.

WSEK1-G20141008-02CoconutEN, 00:04:11.559 - 00:04:13.719

4.4.4.10 Temporal =tawa/=taw

The temporal suffix *=tawa* or *=taw* attaches only to temporal nouns, such as *gnoso* 'now', *kai* '1 day prior/later', *foa* 'later; after'. The semantics of this suffix is not entirely clear, but it often appears to mean 'during that period of time', where the time frame is provided by the host noun. Temporal *=tawa/=taw* is not a durative suffix, but specifies that the predicate of the discourse occurred within some time frame. In Nama there are two morphemes that behave in similar ways: *-taw* is described as 'during' or 'for a period of time', while *-tawa* is described as meaning 'since' (Siegel pers. comm.). It is unclear whether Nmbo has both *=taw* and *=tawa* with distinct semantics, but both forms have been observed.

(97) a. nne mna t\aneh/an, zite=**tawa**.

food.ABS DEM.fw 1sg.A>3pl.U:ipfv.ypst/cook afternoon=**TEMP**

I was cooking food, during the afternoon.

WSEK1-B20170720-02Yayam01YPst, 00:01:00.116 - 00:01:02.37

b. [A public re-enactment of the *mwyang* (bullroarer) in ritual use was performed for documentary purposes. After the mock-demonstration, Smako addressed the young children who had never seen the bullroarer in use.]

bmo totr gs n\äyäre/tate, mwyang däv=**tawa**

2sg.ERG now DEM 2nsg.A:ipfv.npd/hear bull.roarer then=**TEMP**

bä t/maro

3ABS 3sg.U:rmpst/be

You all heard this for the first time, mwyang was like this during that time [of the ancestors]

Speaker smbbSZ

c. [At a village meeting Kawas is making a public announcement to the women involved in planning and running a church gathering to be held in a weeks time.]

Bm bvende lida=va=yo e/m däv=**tawa**.

2ABS 2nsg.POSS leader=COM=ONLY 2nsg:nphd/be then=**TEMP**

You all have got your leaders for the duration of that time [the church fellowship]

WSEK1-B20150726-05PlansKS2,00:01:13.100 - 00:01:15.750

4.4.4.11 Temporal Goal =tio

The semantics of the clitic *=tio* are quite wide, but I propose that it emphasises a particular temporal moment. It occurs on a number of nominal subclasses, and can occur after a variety of other bound morphemes. Temporal goal *=tio* occurs most commonly on locative-marked spatial demonstratives, giving the

forms *yna=no=tio* and *kt=ano=tio*. These two words are respectively translated by Nmbo speakers as 'from here', and 'from there'. The Temporal goal cannot occur on a demonstrative without the locative, **yna=tio*, **kt=tio*.

(98) a. bisi t\iato/tao yna=no=**tio**
arrow 3sg.A>3sg.U:ipfv.rmpst/paint here=LOC=**TEMPG**
y\ake/tai ge n\ovan/
3sg.A>3sg.U:pfv.pret+ipfv/see DEM 3sg:pfv.pst/arrive
"He painted his arrow. Until he saw [the pig] was approaching here!"
WSEK1-B20150924-HuntingWZ, 00:04:54.391 - 00:04:56.663

b. kt ä t\gl/awt yna mrz
DEM.dist 3ABS 3nsg.A>3sg.U:ipfv.rmpst/dance DEM.prox girl
kt=ano=**tio** yägu
DEM.dist=LOC=**TEMPG** morning
There they danced [the dance]. Those girls, until morning...
WSEK1-B20150727-01PastCustomMZ, 00:02:14.316 - 00:02:17.028

The corpus shows a few instances of *=tio* hosted on the noun *srvav*. By itself *srvav* means 'place', but the translation of the bound form *srvav=tio* is given as 'at the same place'.

(99) a. y\ngingo/yn yna srvav=**tio**.
1sg.A>3sg.U:pfv.pret/find DEM.prox place=**TEMPG**
At that same place I had spotted that pig.
WSEK1-G20150902-03PigGS, 00:03:46.180 - 00:03:48.950

b. yna srvav=**tio**... N\unge/nd.
DEM.prox place=**TEMPG** 3du:pfv.pst/stand.up
At the same place... The two of them stood.
WSEK1-G20170629-01Bvk03Cass, 00:03:21.959 - 00:03:24.348

=tio is also hosted on temporal nominals such as *gnoso* 'now' and *nambt* '2~3 days prior/later'. In the examples below, consultants translated *gnoso=taw=tio* as 'until now', and *nambt=tio* as 'recently'.

(100) a. [Dorissa is explaining how she has mastered the craft of basket weaving.]

ämb-ämb-e,	gnoso=taw=**tio**,	yna	är-kitong
some-RDP-DAT	today=TEMP=**TEMPG**	DEM.prox	master

gs	n\amdo/n,	yna	yam	y\avro/tan.
DEM	1sg.A:pfv.pst/become	DEM.prox	activity	1sg.A>3sg.U:ipfv.nphd/do

Now at this stage in my life, as I have become a master, I do this activity and make things all the time. [Original translation by consultant: I've mastered it now, now I'm becoming old, I make this thing all the time.]

WSEK1-G20170707-04DS02MQhm, 00:05:50.469 - 00:05:54.126

There is also one occurrence in the corpus of *=tio* on an already cliticsed noun. In example 101, the noun *kr* is 'death', and is taking the restrictive clitic *=ro* (section 4.4.4.16), which is then followed by *=tio*. From the translation one may infer that the restrictive *=ro* gives a 'only with death/completely dead' interpretation, and the *=tio* provides an additional 'until it was completely dead' interpretation.

(101) a.

k\emna/ngn	de	t\amd/n	kr=ro=**tio**
1sg.A:pfv.fut/think	ALR	1sg:A>3sg.U:pfv.proxin/shoot	death=RST=**TEMPG**

I was thinking then I had already shot [the pig] until it was completely dead...

G20150902-03PigGS, 00:03:58.500 - 00:04:00.600

4.4.4.12 Comitative =va

Nmbo has a comitative morpheme *=va* which expresses accompaniment. It is varyingly realised as *=va*, *=ova*, *=ava*. It typically attaches to nouns (examples

102), but can also occur with spatial demonstratives (102a, second part). While speakers from Bevdvn accept the proximal demonstrative form *yna=va*, they rejected the distal demonstrative form **kt=va*.

In naturalistic speech, there are words that sound like the adverbial demonstrative *gs* (section 4.3.3.2) followed by a *=va* producing the form *gs=va*. Consultants explained these instances as a phonetically reduced form of the particles *bva* and *dva* (section 4.3.1.3).

(102) a.

bés	gane=**va**,	hi	sv-sv,	ynd	yna=**va**.
fire	bark=**COM**	bush.torch	bundle-RDP	1ABS	DEM.prox=**COM**

I was here with the fire bark and the bush torch.

WSEK1-B20150813-05PlaneMZ, 00:01:06.500 - 00:01:09

b.

bm	k\ano/nge	ama	toge-toge=**ava**!
2ABS	2pl.A:pfv.fut/escape	mother.ABS	child-RDP=**COM**

Mother, you all escape with the children!

WSEK1-B20150813-05PlaneMZ, 00:00:48 - 00:00:50

c.

Toka	nq=**ova**	k-n\ango/tao
Toka	anger=**COM**	3sg.A:ipfv.rmpst:ven/return

Toka came back angry. [lit. Toka returned with anger.]

WSEK1-G20150921-LostZS, 00:04:44 - 00:04:45

4.4.4.13 Privative =vna(r)

The privative clitic expresses the absence or lack of the host noun referent. The host may be a concrete object as in example 103a, or may be an abstract concept or emotions such as in example 103b. The proximate spatial demonstratives *yna* can take this clitic, but the distal one cannot, i.e. *yna=vnär*, **kt=vnär*. The clitic is realised in many ways, with *=ovna(r)*, *=avna(r)*, *=vnä(r)*, and *=ävnä(r)* all having been observed.

(103) a. t\mdo/nga de mora-mora=**vna** w/m

3sg.A>3sg.U:pfv.fut/tell ALR medicine-RDP=**PRIV** 1sg.U:nphd/be

She told him like this, "I have no medicine."

WSEK1-B20150909-MissingMoneyMS, 00:04:43.500 - 00:04:45.200

b. bä mnde=**vna** ädi n-ng\amndo/n

3ABS sweet=**PRIV** EMPH 1sg:pfv.pst:and/think

I lost interest. [lit. I thought without desire]

WSEK1-G20151013-10SpiritRG, 00:02:03.755 - 00:02:05.540

4.4.4.14 Similative =nit

The similative *=nit* is a clitic with the meaning 'resembling X', where X is the host nominal (example 104). This morpheme binds to common nouns, and also occurs on manner demonstratives (example 105). I understand the cliticised manner demonstrative to mean 'it looks as though' or 'the situation resembles...'

(104) [Zoga is trying to describe to his audience what an anchor is like.]

a. krta rokar yna-ma t-ng/maro dena...

heavy thing DEM.prox-DEM.ABL 3pl+.U:rmpst/be DEM.mn

karv-bwe=**nit** t-ng/maro

oven-stone=**SIM** 3pl+.U:rmpst/be

These heavy things were like... they resembled oven stones.

WSEK1-G20150826-03AusWorkZG, 00:10:30.136 - 00:10:36.600

(105) [Bévék is participating in a picture task where he has been asked to describe what he sees in the picture in front of him.]

a. OK yna banban=an gym... dena=**nit**

DISC DEM.prox picture=LOC TOPIC DEM.mn=**SIM**

y\aket/an de qévt-an n-n\tor/end

1sg.A>3sg.U:ipfv.nphd/see ALR night=LOC 3du:pfv.pst:ven/exit

Ok in this picture right here... To me it looks like, those two are leaving [the village] at night.

WSEK1-G20170629-01Bvk03Cass, 00:02:45.457 - 00:02:53.719

4.4.4.15 Originative =mn

The originative *=mn* is used to indicate a source relation of a noun. A prototypical example is to express where a person is from, e.g. *Arovwe=mn är*, 'a person from Arovwe' (example 106). The originative can also mark a nominal that is the reason or source of an action. In example 107a the speaker is running away because a helicopter has frightened her, while in example 107b the actions of the pig are explained as being motivated by fear. Indicating the source of an action is also done when *=mn* is bound to the manner demonstrative *mna* (107c). *Mna=mn* is a common way to express what in English is expressed as 'because'.

(106) ynd Arovwe=**mn** mrz w/m

1ABS Arovwe=**ORI** daughter 1sg:nphd/be

I am a woman from Arovwe.

WSEK1-A20140930-CF38RA, 00:03:18.800 - 00:03:20.730

(107) a. yao, ynd helocovta=**mn** n\rs/am.

NEG 1ABS helicopter=**ORI** 1pl.A:ipfv.nphd/run

No, we are running away because of the chopper.

WSEK1-B20150813-05PlaneMZ, 00:02:53.748 - 00:02:55.465

b. mamwi bä de n-ng\rs/at n\ung/äi

pig 3ABS sim 3sg.A:ipfv.nphd/run 3sg.A:pfv.pret/stand.up

kt... end mna=**mn** bérbér=**mn**

DEM.dist road DEM.fw=**ORI** fear=**ORI**

The pig, it's running away, and it stood there... on the road, because of fear.

WSEK1-G20150902-03PigGS, 00:03:22.380 - 00:03:27.600

c. frn-frn=ovnar yäme bä y\räme/tan

pattern-RDP=PRIV mat FUT 1sg.A>3sg.U:ipfv.nphd/make

mna=**mn** ag san qévte y/m.

DEM.fw=**ORI** coconut leaf black 3sg:nphd/be

I will make the mat without patterns because the coconut leaf [which I will use to dye the mat with] is black.

WSEK1-G20141008-02CoconutEN, 00:07:25.308 - 00:07:29.079

Spatial demonstratives marked with the originative *=mn* function as an anaphoric demonstrative, referring back to some preceding state (example 108).

(108) [Rhouda described extensively the migration pattern of peoples to the village of Tais in order to explain how that village came to be dominated by speakers of a language called Nä.]

a. yna=**mn** gnoso dena yamwe... Nä gym

DEM.prox=**ORI** now DEM.mn like.so Nä TOPIC

fronde iy\ätu/tat.

first 3pl.A>3sg.U:ipfv.nphd/speak

That's the reason now... Nä is their main language

WSEK1-A20150923-07TzZiBA, 00:05:26.880 - 00:05:28.578

4.4.4.16 Restrictive =ro

The restrictive *=ro*, or occasionally *=ru*, functions to indicate 'X and nothing more', where X is the host. The restrictive can occur on a variety of nominals including indefinite pronouns (example 110a). The restrictive can occur after a variety of other semantic enclitics. The range of preceding clitics that *=ro* can follow is not known.

(109) a. fivi=**ro** bä y\ne/t gane bä
flesh=**RST** FUT 3sg.A>3sg.U:ipfv.nphd/eat skin FUT
y\wi/t
3sg.A>3sg.U:ipfv.nphd/throw

He will only eat the meat [of the coconut], the skin he will throw it away.

WSEK1-G20150826-07CocoIntZG, 00:03:08.510 - 00:03:09.996

b. yndo smi=**ro** t\nfo/n, t-ng\rs/awn
1SG.ERG tail=**RST** 1sg.A>3sg.U:pfv.proxin/cut 1sg.A>3sg.U:ipfv.rmpst:and/carry
t\aw-iavo/n
1sg.A>3sg.U:pfv.proxin/APP-show

I cut off only the tail, and took it and showed it [to him]...

(G20150826-01CowWG, 00:03:52.926 - 00:03:56.206)

(110) [Examples with a variety of semantic case markers preceding the restrictive.]

a. Dena n\owav/ta. "Snamb=**ru** yz e-nmarengr."
DEM.mn 3sg.A:ipfv.nphd/say how.many=**RST** year.ABS 2sg.U:ipfv.nphd/live

Like this he said: "Just for some years, you will live there."

WSEK1-G20151001-13ThisMorningAN, 00:03:36.736 - 00:03:39.136

b. Nnd=ova=**ro** ädi k\aren/dawm

Grass.skirt=COM=**RST** ITNS 1pl:ipfv.rmpst/wander

We walked around only with grass skirts.

WSEK1-B20150727-02ClothesFT EK, 00:00:17.721 - 00:00:19.331

c. Ynd wrse-h-yan=**ro** kt=ano=tio y\ere/n

1ABS run-INF=LOC=**RST** there=LOC=TEMPG 1sg.A>3sg.U:pfv.pst/catch

I kept running until I caught it.

WSEK1-B20141007-PigYS, 00:05:45.925 - 00:05:47.520

4.4.4.17 Only =yo

The morpheme *=yo* signifies the host having done something by him or herself (examples 111). The only examples in the corpus are of this morpheme attached to human nouns and absolutive pronouns. In all cases the verb is a one-place predicate. Like the restrictive *=ro*, this *alone* clitic can occur after a variety of semantic case markers (112), the full range of which have not been tested for.

(111) a. ynd=**yo** k\awatambne/tan.

1ABS=**ONLY** 1sg:ipfv.rmpst/learn

I learned by myself.

WSEK1-A20150924-12BasketKA, 00:03:34.761 - 00:03:36.601

b. [A child asked his mother a question. When a sibling answered instead, the child retorted to the sibling:]

e, ama=**yo** gea k\owav/tangai

EXCL mother=**ONLY** SUPP 3sg:pfv.fut+ta+pret/speak

Hey, mother should have answered herself!

WSEK1-B20141007-PigYS 00:03:52.743 - 00:03:54.573

(112) [Examples with a variety of semantic case markers preceding and following the restrictive.]

a. bm bvende lida=va=**yo** em däv=tawa.
2ABS 2nsg.POSS leader=COM=**ONLY** 2nsg:nphd/be when=TEMP

You all have got your leaders for the duration of that time [the church fellowship].

WSEK1-B20150726-05PlansKS2,00:01:13.100 - 00:01:15.750

b. [Fangore is talking about how she was exchanged for another girl in marriage for the benefit of her brother Maiwa. Fangore only has one brother.]

tande hakr=**yo**=wt, Maiwa=yot
1sg.POSS brother=**ONLY**=BEN Maiwa=ALL

For my only brother, for Maiwa.

WSEK1-G20170723-02FYCocohm, 00:01:26.389 - 00:01:27.955

4.5 Verbs and their Morphology

This section of the sketch grammar will deal with complex finite verbs and their morphology.

Nmbo has two types of morphologically defined verb complexes; *prefixing verbs*, and *ambifixing verbs* (we will define these terms shortly). These two verb types share most of their architecture, so an overview of the maximal verb complex will be provided in the template section (4.5.1). A morpheme-by-morpheme description of the slots will then be provided, with the understanding that morphosyntactic functions are in fact distributed across multiple morphemes. For example, TAM is distributed across multiple affixes: the u-prefix, the dual/nondual+TAM thematic suffix , and the a-suffix. It is only when all verbal affixes are viewed in relation to each other that the total meaning of the verb complex is made clear; what Carroll (2016) describes as *distributed exponence* for the Tonda branch Yam language Ngkolmpu. The results is a circumfixal paradigm of combined argument marking and TAM. This section will describe how argument marking,

and TAM marking, are achieved by ambifixing verbs in section 4.5.7). This verbal morphology section will end with a description of the prefixing verbs, which are semantically distinct from ambifixing verbs, and have a few morphological peculiarities (section 4.5.8).

The terms prefixing and ambifixing verbs are used by Evans (2012, 2015a) to describe the structure of inflecting verbs in Nen. Döhler (2018) also does so for the Tonda language Komnzo. It is a morphologically defined distinction of how argument structure is expressed; predominantly by prefixation, or by prefixation and suffixation. When referring to the verbal indexation of arguments I use the terms *actor* and *undergoer* as is convention for the Yam languages, where actor includes A of transitives and S of dynamic intransitives, and undergoer includes P, R, and A of "statives" (Evans (2015a):544).[9] Table 4.21 shows the loci of argument indexing on the different verb types.

Prefixing verbs, which are always monovalent, cross-reference the person, and part of the number of the sole argument by prefix. Prefixing verbs can be grouped further into two types: the positional verbs with unique suffixes that mark number, and verbs that are built up from the copula. Prefixing verbs can morphologically express four different TAMs: the non-prehodiernal (NPHD), yesterday past (YPST), primordial past (PRIM), and the remote past (RMPST). These are the TAM categories expressed for verbs with imperfective aspect (section 4.7).

Ambifixing verbs can mark two arguments through prefixation and suffixation. Ambifixing verbs can be monovalent, divalent, and trivalent. Divalent verbs index the undergoer by prefixation, and actor by suffixation. Monovalent verbs index the sole argument on the actor marking suffix (the a-suffix), while the prefix is a person/number invariant middle marker. The marking of the third argument in a trivalent verb is described in the section on three-place predicates (section 4.8.3).

4.5.1 Morphemes of the Verb Template

Prefixing and ambifixing verbs share much of their architecture. In this section I will analyse separately the slots of a maximal ambifixing verb template, with the

[9] As Evans (2015c) notes, the S of verbs of motion 'come' and 'go' are agentive, and the verbs are not technically stative. I use scare quotes to indicate that while 'come' and 'go' are semantically dynamic, they are derived from the copula which can be characterised as stative.

	Monovalent		Divalent
	(stative)	(dynamic)	
	Prefixing	Ambifixing	Ambifixing
Examples	**w**-m	nowav-**tan**	**y**-mde-**tan**
	1sg.U-be	talk-**1sg.A**	**3gs.U**-speak-**1sg.A**
	"I am."	"I am talking."	"I am speaking to him."
	w-ng-m	nrs-**at**	**n**-mi-**tan**
	1sg.U-AND-go	run-**1sg.A**	**2sg.U**-ask-**1sg.A**
	"I am going."	"I am running."	"I am asking you."
	w-akiongr	naho-**tan**	**ta**-gme-**tan**
	1sg.U-stand	dress-**1sg.A**	**3sg.U**-hit-**1sg.A**
	"I am standing."	"I am dressing."	"I am hitting them."

Table 4.21: Examples of argument marking by prefixation and/or suffixation. All examples are in the non-prehodiernal TAM.

caveat that these individuated analysis must in the end be viewed in relation to one another to specify the full meaning expressed by the entire verb complex. A unified view of the template presents ambifixing verbs as forming a circumfixal paradigm (section 4.5.7). The particularities of prefixing verb suffixes will be described in section 4.5.8).

Table 4.23 shows the maximally abstracted ambifixing verb template, with everything left of the root (i.e. including the diathetic) being applicable to prefixing verbs:

Ambifixing Template						
Prefixing Template						
Prefix		Stem		Desinence		
U-PREFIX	(DIR)	DIATH	ROOT	θ	(θ EXT)	A-SUFFIX
Undergoer person and number	Motion hither and thither	Valency increasing and decreasing	Root	Dual and Non-Dual number, Macro aspect	Micro TAM	Actor person and number

Table 4.22: Nmbo verbal template. DIR = directional slot; DIATH = diathetic; θ = thematic; θ+EXT = thematic extension

The verb template can be broken up into three parts: the *stem*, *prefix*, and *desinence* (Evans 2015a).

The stem consists of the *root* and an optional *diathetic prefix*. The root provides the basic lexical meaning, while the diathetic is responsible for valency increasing and decreasing functions. Broadly speaking, valency increasing prefixes take the form *w(a)-*, while valency decreasing prefixes begin with *a-* or *ä-*.

The greater prefix consists of an obligatory undergoer marking prefix, the *u-prefix*, and an optional *directional* prefix. As the name suggests, the u-prefix indexes the person and number of the undergoer-like argument (P). It also encodes TAM information. There are two sets of u-prefixes, labelled α and β because the semantics of the sets by themselves are not transparent (see table 4.24 for forms). The overall verbal TAM, and precise number of the undergoer is only clear when the u-prefix is taken together with the form of the desinence. The directional prefix provides motion details. Prototypically this is either 'towards something, e.g. the narrator, a location' (*venitive* 'hither'), or 'away from something. e.g. the narrator, a location' (*andative* 'thither").

The desinence is comprised of a *thematic* suffix (θ), a *thematic extension* (θ+EXT), and an actor marking suffix (*a-suffix*).

The thematic and thematic extensions provide TAM specifications, but the thematic also marks number, namely dual vs. non-dual number. The dual/non-dual thematic form is sensitive to the aspect of the verb. Nmbo verbs are morphologically categorised as either imperfective or perfective. When the aspect is imperfective the thematic makes an alternation of *-w*(dual)/*-ta*(non-dual) or ø/*-ta*. The *-w* form of the dual occurs when the actor is 1st or 3rd person, and it is the øwhen the actor is the 2nd person (table 4.27). If the verbal aspect is perfective, the alternation is *-e*(dual)/ø(non-dual) (examples 114). The series of the u-prefix and the MA indicated by the thematic produces an overall predicate TAM. These predicate TAMs can be extended further by the thematic extension (section 4.5.5).

The a-suffix marks the person and number of the actor argument, and has different forms depending on the verbal TAM. First person forms are consistently *-n* (1sg) and *-m* (1nsg) across the different verbal TAMs, but the second and third persons show more variety.

(113) Imperfective verb *mih* ‘ask’ with *-ta*(non-dual)/*-w*(dual) alternation.

a. [Singular actor, singular undergoer]

Yndo	bä	y-mi-**ta**-n.
1sg.ERG	3ABS	3sg.U:a-do-**ipfv.ndu**-1sg.A

I am asking her.

b. [Singular actor, dual undergoer]

Yndo	bä	sombwi	e-mi-**w**-n
1sg.ERG	3ABS	two	3nsg.U:ɑ-ask-**ipfv.du**-1nsg.A

I am asking those two.

c. [Dual actor, singular undergoer]

Yndvem	är	sombwi	bä	ämbru	y-mi-**w**-m.
1nsg.ERG	person	two	3ABS	one	3sg.U:ɑ-ask-**ipfv.du**-1nsg.A

The two of us are asking her.

d. [Plural actor, plural undergoer]

Yndvem	är	nambi	bä	nambi	e-mi-**ta**-m
1nsg.ERG	person	three	3ABS	three	3nsg.U:ɑ-ask-**ipfv.ndu**-1nsg.A

We three are asking those three.

Fieldnotes 2016

(114) Perfective instance of verb *farh* ‘carve’ with *ø* zero (non-dual)/*-e*(dual) alternation.

a. [Singular actor, singular undergoer]

Yndo	bä	yvarn
		y-var-**ø**-n
1sg.ERG	3ABS	3sg.U:ɑ-carve-**pfv.ndu**-1sg.A

I carved it.

b. [Singular actor, dual undergoer]

Yndo	bä	sombwi	ewvar**e**n
			ew-var-**e**-n
1sg.ERG	3ABS	two	3nsg.U:ɑ-carve-**pfv.du**-1nsg.A

I carved two things.

c. [Dual actor, singular undergoer]

Yndvem	är	sombwi	bä	ämbru	yvar**e**m.
					y-var-**e**-m.
1nsg.ERG	person	two	3ABS	one	3sg.U:ɑ-carve-**ipfv.du**-1nsg.A

The two of us carved one thing.

d. [Plural actor, plural undergoer]

Yndvem	är	nambi	bä	nambi	ewvarm
					ew-var-**ø**-m
1nsg.ERG	person	three	3ABS	three	3nsg.U:ɑ-carve-**ipfv.ndu**-1nsg.A

We three carved three things.

Fieldnotes 2016

As this brief extrapolation of the template demonstrates, the Nmbo verb com-

plex can be analysed as a template with morpheme slots, but the information of arguments and TAM is scattered across various slots which are usually non-contiguous. Nmbo's verb complex demonstrates what is called *distributed exponence* (Caballero and Harris 2012). Distributed exponence in Ngkolmpu is described by Carroll (2016) as a phenomenon where the realisation of individual feature values (e.g. agreement, TAM) is complex with involvement of multiple affixal elements. Each morpheme is required in order to provide a precise interpretation of the inflected verb (Carroll 2016:172).

In the following sections we will go through each morpheme in more detail than the overview provided above, with attested forms, and examples of their particular functions. This will demonstrate how the morphemes combine to produce an overall inflected meaning.

4.5.2 Stem: Root and Diathetic Prefix

The stem provides the basic semantics of the predicate, and lays the foundations for morphological processes to build up the verb complex. The stem consists of the *root*, and a *diathetic prefix* which changes valency.

Nmbo has a phonotactic particularity where vowel initial verb roots are always monovalent. Consonant initial roots are di- or trivalent. The diathetic prefix changes the valency frame, and does so in a way that is consistent with the phonotactic constraint: diathetic *w-* prefixed to a monovalent root derives a benefactive/applicative stem, while a vowel V- prefixed to a di/trivalent root derives an anticausative, antipassive, or reflexive/reciprocal stem. The benefactive/applicative diathetic indicates the existence of a third R argument required by the verb.

The vowel-form of the valency decreasing prefix varies between roots, with *a* [ɑ] , *ä* [æ] , *á* [ɐ] having all been observed. There is a tendency for the diathetic vowel to harmonise with the first vowel of the root if the vowel is [ɑ] or [æ] (e.g *wanoh* 'to wake someone up' > *a-wanoh* 'to awake', *gärngh* 'to turn something over' > *ä-gärngh* 'to turn over by one's self'), but this is not always the case (e.g. *werh* 'to hold something' > *ä-werh* 'to hold on').

The diathetic is best analysed as part of the stem for two reasons. Firstly, the lexical meaning of the stem changes in a transparent way suggesting that the diathetic is a productive valency alternating mechanism, i.e. the valency

Root	Gloss	Derived Root	Gloss
Valency Increasing: Monovalent to Di/Trivalent			
armboh	to ascend	*w-armboh*	to carry something up
asrh	to pour out by self	*wasrh*	to pour something out
ämsuh	to sit	*w-msuh*	to put something down
ámbrtoh	to die out (i.e. a flame)	*w-mbrtoh*	to put out a fire or light
umengh	to gather	*w-umengh*	to gather people
Valency Decreasing: Divalent to Monovalent			
ferh	to undo something	*ä-verh*	to become undone
frengh	to prepare something	*ä-vrengh*	to prepare oneself
gärngh	to flip/turn something over	*ä-gärngh*	to turn around
nengh	to uncover something	*a-nengh*	to become uncovered
lih	to swallow something	*á-lih*	to swallow
soh	to scratch something	*a-soh*	to scratch oneself
waingh	to overtake someone	*a-waingh*	to pass by
wanoh	to wake someone up	*a-wanoh*	to wake up
werh	to hold something	*ä-werh*	to hold on

Table 4.23: Diathetic prefix and derived stems

differences are not lexical (e.g. *ferh* 'to undo something' > *ä-verh* 'to become undone', *dñih* 'to hide oneself' > *ä-dñih* 'to hide something').

Secondly, phasal verb constructions (section 4.9.1) involving nonfinite verbs can have base root and derived stem forms as the lexical auxiliary. The inflecting phasal verb then matches the valency of the nonfinite verb. In example 115a the non-finite *soh* 'to scratch something' requires two arguments, and the phasal verb *wavngoh* 'to try and do something' indexes two arguments accordingly, i.e. *y-avngo-yn*. Example 115b with the non-finite *a-soh* 'to scrach oneself' has a middle form of the phasal verb *ovngoh*, which indexes the sole argument on the a-suffix, i.e. *n-ovngo-yn*:

(115) Contrast of *soh* 'to scratch something' (a) and *asoh* 'to scratch self' (b):

a. ynd **so**-h **y**-avngo-yn

1ABS scratch-INF **3sg.U**-do-1sg.A(pfv.pret)

I was trying to scratch him.

b. ynd **aso**-h **n**-ovngo-yn

1ABS scratch.self-INF **M:α**-do-1sg.A(pfv.pret)

I was trying to scratch myself.

Fieldnotes 2017 Book 1:10

The description of valency alternation types in terms of the verbal morphology and nominal case marking will be covered in the clauses and valency section of the sketch grammar (section 4.8).

4.5.3 Undergoer Prefix (u-prefix)

The undergoer prefix (*u-prefix*) indexes the person and number of undergoer arguments, and contributes to TAM specifications. For prefixing verbs the u-prefix indexes the sole argument (example 116). For ambifixing verbs it indexes the undergoer argument (117a). For monovalent ambifixing verbs with no undergoer argument, the u-prefix takes a person and number invariant *middle* form, and the sole argument is indexed on the a-suffix (117b).

In three-place predicates, the u-prefix can index the recipient argument R (118a), or the object of transfer T (118b). It appears to be more common for R to be indexed on the u-prefix rather than T. For case marking on the free nominals R and T, see section on trivalent constructions (4.8.3)

(116) [U-prefix on prefixing verbs.]

a. [The suffix on the positional verb *-ngrwn* is a non-dual remote past suffix]

Bä kt **t**/mare-ngrwn.

3ABS DEM.dist **3sg.U**:rmpst/wait

He was waiting there.

WSEK1-G20150902-03PigGS, 00:17:45.848 - 00:17:47.348

b. Bä **ta**/mare-ngrwn.

3ABS **3pl.U**:rmpst/wait

They were waiting.

WSEK1-G20150902-03PigGS, 00:17:48.673 - 00:17:49.941

(117) [U-prefix behaviour on ambifixing verbs: Example (a) indexing the undergoer of a divalent verb, (b) as a middle marker in the same slot.]

a. yndo ag **y**-nfo-n.

1SG.ERG coconut.ABS **3sg.U**-cut-1sg.A(pfv.past)

I just cut the coconut.

Fieldm notes 2014:

b. ynd **n**-wi-n.

1ABS **M:ɑ**

I just fell.

Field notes 2017:

(118) [U-prefix behaviour on trivalent ambifixing verbs.]

a. [Indexing the recipient of a divalent verb.]

be=wt ä **k**-n-a-ramo-n

2sg.DAT=BEN FUT **2sg.U**-VEN-BEN/APP-give-1sg.A(pfv.proxin)

nu=ama wrse-här rokar

water=PERL run-AG.NMZR object

I will give you a dinghy.

WSEK1-B20150928-MPVisitLS, 00:04:22.755 - 00:04:26.360

b. [Indexing the theme of a divalent verb.]

Ymo ämbru yn̄g̈ **y**-ramo-i ta-vav.

3sg.ERG one bag.ABS **3sg.U**-give-3sg.A(pfv.pret) **1sg.DAT=GOAL**

He gave me one bag.

Fieldnotes 2017 Book2, p.9

The u-prefix makes person distinctions of first, second, and third persons, and number distinctions along a singular vs. non-singular division. Each person-number configuration has two forms (table 4.24), forming two sets of prefixes series. As has been noted for Nen (Evans 2015c:1076) and Nama (Siegel 2014b:214), the u-prefixes do not carry specific TAM meanings by themselves, but aids in the specification of TAM once combined with the the thematic (θ), thematic extension (θ+EXT), and a-suffix. The convention of Yam language description is to thus err on the side of caution and use the semantically neutral labels α and β when describing the different u-prefix sets.

In examples 119, all affixes are held constant save the u-prefix, and only the alternation of the prefix series provides the differing TAM interpretation. The examples here include adverbials for explicit clarification, but the meaning would be communicated to speakers without these.

(119) a. Gnoso **y**-nfwe-tan

now **3sg.U:α**-cut-1sg.A(ipfv.nphd)

Now I'm cutting it.

	ɑ	β
1SG	w-	q-
1NSG	yn-	tn-
2SG	n-	k-
2NSG	e(w)-	ta-
3SG	y-/ i*-	t-
3NSG	e(w)-	ta-
middle	n-	k-

Table 4.24: Nmbo undergoer prefix forms by series.
* = when the verb root begins with [i, j], the ɑ-set 3sg prefix may be realised in this form. e.g. *yeh* 'to plant' → *iyeyn* 'I planted it'.

b. Kai **t**-nfwe-tan

± 1day **3sg.U:β**-cut-1sg.A(ipfv.nphd)

Yesterday I was cutting it.

Fieldnotes 2014 Book of Collections:8

4.5.4 Directional Prefix

This optional prefix provides motion-based spatial-directional information. There are two kinds: the form *n-* prototypically signals motion towards the speaker (*venitive* 'hither'), while *ng-* signals motion away (*andative* 'thither'). The clearest manifestation of these directionals can be seen in the construction of 'come' and 'go' from the copula. The root form $\sqrt{m}$[10] 'to be' is inflected with the prefix indexing the person and number of the sole argument, and the directional is prefixed immediately left of the root (examples 4.5.4).

(120) Contrastive examples of directional prefix.

a. y-**n**-m

3sg.U:ɑ-**VEN**-be

S/he came.

[10]This notation indicates that the root form is not a free word, unlike the infinitive form of other lexical verbs.

b. y-**ng**-m

3sg.U:ɑ-**AND**-be

S/he went.

These directional prefixes can extend motion into other domains. For verbs of motion such as *ovarh* 'to arrive' the directionals provide additional locational information.

(121) a. yna=no n-n\ovar/st

DEM.prox=LOC 3pl.A:pfv.pst:**ven**/arrive

They arrived here.

WSEK1-B20150805-02NokiaYZ, 02:21.274 - 02:23.800

b. kt=ano n-**ng**\ovar/ym

DEM.dist=LOC 1pl.A:pfv.pret:**and**/arrive

We arrived there.

WSEK1-G20150826-07CocoIntZG, 00:03:29 - 00:03:36

An interesting example comes from a narrative, where the alternation of this directional suffix is interpreted as negation. The andative signals 'I give', while the venitive a 'I will not give' meaning. Nothing else in either clause is likely to be providing this negation. It may be that the 'thither' is providing a 'giving away from speaker' semantics, while the 'hither' is a 'keeping here with the speaker' semantics.

(122) a. k-n\m aweh, mé k-**ng**\a-ramo/n.

come(IMP) DISC CONT 1sg.A>2sg.U:pfv.proxin:**and**/APP-give

Come here, I'll give you it.

WSEK1-B20150727-01PastCustomMZ, 01:35.576 - 01:36.820

b. bä k-**n**\a-ramo/n.

FUT 1sg.A>2sg.U:pfv.proxin:**ven**/APP-give

'I won't give you it. [lit. I will give towards me.]

WSEK1-B20150727-01PastCustomMZ, 01:38.450 - 01:39.536

An unusual function found in some Nambu languages is the appropriation of the andative *-ng* to construct large plural number (Evans, 2019c). Nmbo is one of the languages that does so. Ambifixing transitive verbs use this morpheme to indicate that the undergoer is a large plural, which is distinct from a regular plural. This must be done, however, with the u-prefix in the singular number.[11]

Table 4.25 shows the construction of large plurals in ambifixing verbs, but this kind of construction also occurs for prefixing verbs. This is covered separately in the prefixing verb section 4.5.8.

Ambifixing Verb Example			
Inflected Verb Form	u-prefix	dual+a-suffix	Translation
y-ävätutan	3sg	ndu+1sg	'I am shooting it'
ew-ävätuwn	3nsg	du+1sg	'I am shooting two things'
ew-ävätutan	3nsg	ndu+1sg	'I am shooting several things'
y-**ng**-ävätutan	3sg	ndu+1sg	'I am shooting many many things'

Table 4.25: The construction of large plural from the imperfective verb *wävätuh* 'to shoot something'. Fieldnotes 2016 Book 3: 16

4.5.5 Thematics (Thematic, Thematic Extension, and Secondary Thematic)

The thematic suffix to the right of the root indicates dual number and TAM. It expresses a dual vs non-dual number distinction of either actor and/or undergoer argument. The form is contingent on the aspect of the verb (summarised in table 4.26), as well as on further TAM specifications made by the thematic extension. I will describe the thematic and thematic extensions separately for the time being, given that the thematic in particular exhibits a clear pattern of forms.

[11] Evans (2019c) notes that in Nen, if perfective verbs are to construct a large plural in this manner, the non-dual thematic morpheme must take the form *-ta*, which is the imperfective non-dual form (p.22). This has not been tested for in Nmbo.

As mentioned earlier (section 4.2.2), Nmbo verbs can be grouped into aspectual groups: imperfective and perfective (more details in 4.7). If the verb is imperfective, the thematic slot makes an alternation of *-ta*(non-dual) with some other form. Dual is realised as *-w* (first and third person duals) or ø(second person duals) when TAM is the *non-prehodiernal*, or *yesterday past*. Dual is ø when in the *remote past*, or *primordial past*. The semantics of these TAM labels are described in section 4.7.

The non-dual imperfective *-ta* undergoes morphophonemic changes depending on the final consonant of the verb root. There are not hard and fast rules, but general tendencies. For example, the *-ta* is realised as *-da* when the verb root ends in [ᵑg], i.e. the verb root *umeng* 'to congregate' becomes numen-**d**am 'We are congregating'. Below are some known transformations.

- r + ta → *na* (example 123a)
- ᵑg+ ta → *da* (example 123b)
- n + ta → *na* (example 123c)
- s + ta → *sa* (example 123d)
- h + ta → *ha* (example 123e)

(123) a. r → n

Ynd	anu	numbe**na**n
		n-umbar-**ta**-n
1ABS	bath	M:ɑ-bathe-**ipfv.ndu**-1sg.A

I am bathing.

Nmbo Dictionary 2018, entry for *umbärh* 'to bathe'

b. ᵑg→ d

Äusa-äusa	Dimsisi=n	kumen**da**t.
		k-umeng-**ta**-t
women	Dimisisi=LOC	M:β-gather-**ipfv.ndu**-3nsg.A

The women congregated at Dimsisi.

Nmbo Dictionary 2018, entry for *umengh* 'to gather'

c. n → n

Mwingane	yalnan
	y-walan-**ta**-n
yam.type	3sg.u-scrape-**ipfv.ndu**-1sg.A

I am scraping the Mwingane yam.

Nmbo Dictionary 2018, entry for *walanh* 'to scrape'

d. s → s

Nu	yrosat.
	y-rosa-**ta**-t
water	3sg.U-search-**ipfv.ndu**-3nsg.A

They are searching for water.

WSEK1-A20150917-1SinkingBA, 00:06:12.577 - 00:06:13.618

e. h → h

Yndo	yna	dram	yrohan.
			y-rohé-**ta**-n
1sg.ERG	this	drum	3sg.u-drag-**ipfv.ndu**-1sg.A

I am dragging the drum.

WSEK1-A20150917-1SinkingBA, 00:07:12.738 - 00:07:14.520

Perfective verbs also show non-dual/dual alternations in the thematic. For *perfective past* and *proximate inceptive*, the alternation is *-s*(non-dual)*-e*(dual), but when the actor is singular, the non-dual thematic can take the zero form ø. The *perfective preterite* also has a consistent *ø*(non-dual)*-e*(dual) alternation. The *perfective future* paradigm can be analysed as making a *-ng*(non-dual)*-e*(dual) alternation. Table 4.26 summarises these thematic alternations, while table 4.27 realises these abstraction in context of select verbs.

	Dual	Non-Dual	TAM	Example: Dual	Example: Non-Dual
Imperfective	-w*	-ta	NPHD	narmbo-w-m	narmbo-ta-m
			YPST	karmbo-w-m	karmbo-ta-m
	ø	-ta	PRIM	narmbo-ø-mm	narmbo-ta-mm
			RMPST	karmbo-ø-wm	narmbo-ta-wm
Perfective	-e	-s**	PST	narmb-e-m	narmbo-s-m
			PROXIN	karmb-e-m	karmbo-s-m
	-e	ø	PST	narmb-e-m	narmbo-ø-m
			PROXIN	karmb-e-m	karmbo-ø-m
			PRET	narmb-e-aym	narmbo-ø-ym
	-e	-ng	FUT	karmb-e-m	karmbo-ng-m
			PHAB	karmb-e-aym	karmbo-ng-aym

Table 4.26: Thematic dual alternation patterns. Verb root is *armboh* 'to ascend'. * = first and third person duals only ; ** = first person only

Actor	1DU	1PL	2DU	2PL	3DU	3PL
Imperfective 1: Non-prehodiernal and Yesterday Past						
armboh 'to ascend'	narmbo-**w**-m	narmbo-**ta**-m	narmbo-**ø**-te	narmbo-**ta**-te	narmbo-**w**-t	narmbo-**ta**-t
mih 'to ask'	t-mi-**w**-m	tmi-**ta**-m	tmi-**øte**	t-mi-**ta**-te	t-mi-**w**-t	t-mi-**ta**-t
Imperfective 2: Primordial and Remote Past						
mih 'to ask'	y-mi-**ø**-mm	ymi-**ta**-mm	y-mi-**ø**-mt	y-mi-**ta**-m	y-mi-**ø**-mnd	y-mi-**ta**-mnd
	t-mi-**ø**-wm	tmi-**ta**-wm	t-mi-**ø**-wt	t-mi-**ta**-wte	t-mi-**ø**-t	t-mi-**ta**-wt
Perfective 1: Perfective Past						
oroh 'to leave alone'	nor-**e**-m	noro-**s**-m	nor-**e**-nd	nor-**e**-nd	nor-**e**-nd	noro-**s**-t
Perfective 2: Perfective Preterite						
wavroh 'to do'	yavro-**e**-aym	yavro-**ø**-ym	yavo-**e**-aende	yavro-**ø**-ynde	yavro-**e**-aend	yavro-**ø**-ynd
Perfective 3: Perfective Future						
amdoh 'to become'	kamd-**e**-m	kamdo-**ng**-m	kamd-**e**-nd	kamdo-**ng**-e	kamd-**e**-nd	kamdo-**ng**-e

Table 4.27: Example partial paradigm showing realisation of imperfective dual thematic by first, second, and third person agent. For the purposes of exposition, the thematic extension and a-suffix have been grouped together. Where there is an undergoer, it is held constant at 3SG.

The thematic extension slot occurs to the right of the thematic and provides additional TAM information. An extension of *-m* left of the imperfective thematic builds a combined non-dual form *-tam* for the *primordial* TAM. A *-w* left of the imperfective thematic likewise build the form *-taw* for a non-dual argument *with remote past* TAM.

The perfective has a single thematic extension *ay*, which seems to signal a distant past. Siegel (2014b) calls the corresponding form in Nama the "perfective remote" suffix (p.225). For example an extension *-ay* left of the perfective thematic creates a combined form *-eay* for a dual argument in the preterite, e.g. *narmb-e-m* 'Two of us went up' vs *narmb-e-ay-m* 'Two of us went up a long time ago'. The *ay* often undergoes a morphophonemic process where the initial [a] of the suffix is dropped ($ay \rightarrow y$). The conditioning factors for this change have not been systematically explored, but it seems to occur when the verb root ends in a vowel, and the actor is non-dual, e.g. *oroh* 'move', n-oro-ay-n → *noroyn* 'I moved'; *ämsuh* 'to sit', n-ämsu-ay-n → *nämsuyn* 'I sat.'

While I have described the thematic and thematic extension separately, the combined forms are the most informative units in terms of TAM and agreement. Table 4.28 shows a paradigm of the combined thematics (θ and θ+EXT). With this view we can see the formal grouping of imperfective verbs as 'verbs where the thematics contains *-ta* in a non-dual form', and perfective can be groups as 'verbs where the thematics contains *-e* in a dual form'.

	Imperfective				Perfective				
	ipfv.imp	nphd/ypst	prim	rmpst	pfv.imp	past/proxim	pret	fut	phab
non-dual	ta	ta	tam	taw	ø	s**	(a)y	ng	ngay
dual	ø	w*	m	w	e	e	eay	e	eay

Table 4.28: Subparadigm of thematic (θ+ θ+EXT).
* = first and third person duals only ; ** = first person only

A final complication is that of a *secondary thematic* slot. Evans posits this slot for Nen, (forthcoming pp.86-87), where some verbs insert a basic non-dual thematic *-ta* between the verb stem and suffix. The result is a verb with modified lexical aspect, such that a perfective verb is coerced into an imperfective (i.e. continuative, iterative, durative) interpretation. This also happens in Nmbo. For example a verb inflecting for the perfective past habitual can take a secondary thematic *-ta*, e.g. kowav-ta-ngay-n. This secondary thematic does not have an

alternate dual form, which forms the basis of why it should be distinguished from the imperfective thematic proper.

In addition to *-ta*, Nmbo arguably has a *-tang* secondary thematic. It occurs immediately to the right of the verb stem when present. It is analysed as a form distinct from the *-ta* because it too exhibits no sensitivity to dual number. For example the elicited form *tavrotangayn* which was translated as 'I should have done it', could be segmentally analysed as perfective past habitual *tavro-ta-ngayn*, but the dual actor form in this paradigm is given as *tavro-tang-e-aym*. If this were a case of the past habitual extended by *-ta*, the expected form should be *tavro-ta-eaym* or some other phonologically licit form.

The semantics provided by the *-tang* are not entirely clear. Some speakers offered that the "size of the activity" is relevant, although I am unsure what this may mean. Unfortunately I have not been able to satisfactorily interrogate the behaviour or semantics of this particular paradigm, but perhaps it is due to the morphological and semantic complexity that some speakers exhibit variability in the combinatorics of the prefix set. One speaker consistently produced past habitual forms extended with -tang with the ɑ-set of prefixes rather than the β-set. Tantalisingly, this speaker is a daily Nen-Nmbo bilingual woman.

Actor	Past Habitual TAM suffix	+ tang suffix	+tang example *owavh* 'to speak'
1sg	-ng-ay-n	-tang-ø-ay-n	*kowavtangayn*
1du	-e-ay-m	-tang-e-ay-m	*kowavtangeaym*
1pl	-ng-ay-m	-tang-ø-ay-m	*kowavtangaym*
2sg	-ng-ay-ø	-tang-ø-ay-ø	*kowavtangay*
2du	-e-ay-ng	-tang-e-ae-nde	*kowavtangeaende*
2pl	-ng-ay-nge	-tang-ø-ay-nde	*kowavangeynde*
3sg	-ng-ay-ø	-tang-e-ay-ø	*kowavtangeay*
3du	-e-ay-ng	-tang-e-ae-nd	*kowavtangeaend*
3pl	-ng-ay-ng	-tang-ø-ay-nd	*kowavtangaynd*

Table 4.29: Subparadigm for suffix with secondary thematic *-tang*.

4.5.6 Actor Suffix (a-suffix)

The a-suffix is the right-most element of an ambifixing verb template. The function of the a-suffix is to mark the person and number of the actor argument.

This can be the sole argument of a monovalent verb, or the actor of a di- or trivalent verb. Like the u-prefix, the a-suffix makes first, second, and third person distinctions, and a singular/non-singular number distinction. Unlike the u-prefix, however, the a-suffix has a wider variety of forms for each possible person/number combination.

The form of the first person a-suffixes are consistent across various TAMs: *-n* for singular and *-m* for non-singular. Second and third persons show relatively consistent patterns, but the forms vary depending on the verbal aspect. Second and third singular a-affix cells in any given paradigm are syncretic. Second and third non-singular suffixes often resemble each other in form, where the 2nd non-singular is formed by adding an *-e* to the 3rd non-singular form (e.g. third non-singular imperfective remote past *-tawt*, second non-singular imperfective remote past *-tawte*, third non-singular perfective preterite *-yng*, second non-singular perfective preterite *-ynge*).

There are also variations in some of the cells. The form for the third singular imperfective primordial has variously been *-tam* or *-tamng*, and the third singular perfective preterite similarly varies between *-y* and *-yng*. Overall, it seems that the a-suffixes of the imperfective paradigm show a lot of consistency in the speech community, but the perfective forms are more variable are require a more robust investigation.

Much like the thematic extension, the a-suffix has many exceptions in forms. As mentioned in the discussion about the thematic extension, the a-suffix is best viewed in conjunction with the thematic and extension rather than as an independent morpheme. Table 4.30 shows known forms for the desinence which is comprised of the thematic (+ extension) and a-suffix.

4.5.7 Ambifixing verbs and the circumfixal paradigm

Given that multiple realisations of various exponences in the verb complex (e.g. imperfective dual thematic can be *-w* or ø), we can see that an an over-specified morpheme based analysis show its limits. A *word-and-paradigm* based approach (Anderson 1992, Blevins, Milin, and Ramscar 2015) provides a more elegant view where the u-prefix and desinence are treated as forming a *circumfixal paradigm* (Evans 2015a:544, 2015b:1075). In such a view the prefixes and suffixes specify person/number agreement and TAM when viewed in conjunction to one another
.

	Imperfective			Perfective			
	NPHD, YPST	PRIM	RMPST	PST, PROXIN	PRET	FUT	PHAB
1sg	-tan	-tamn	-tawn	n /-sn	-yn	-ngn	-ngayn
1du	-wm	-mm	-wm	-em	-eaym	-em	-eaym
1pl	-tam	-tamm	-tawm	-sm	-ym	-ng-m	-ngaym
2sg	-t	-tam	-taw	-sa /ø	-y	-nga	-ngay -ngayng
2du	-te	-mnde	-wt	-end	-ea-ende	-end	-eaynd -eayng
2pl	-tate	-tamnde	-tawte	-s-te	-ynde	-nge	-ngaynde -eaynge
3sg	-t	-tam -tamng	-taw	-sa	-y -yng	-nga	-ngay -ngayng
3du	-wt	-mnd	-t	-end	-eaend	-end	-eaynd -eayng
3pl	-tat	-tamnd	-tawt	-st	-ynd	-ng	-ngaynd -ngayng

Table 4.30: Observed desinence forms (thematic, thematic extension, and a-suffix).

For example if we take the verb form *kmitawn*, we know from the u-prefix is from the β-set, so the verbal TAM must be one of the following: imperfective yesterday past or remote past, or perfective proximate inceptive, future, or past habitual. The prefix form *k-* could be either a middle marker or a second person singular undergoer, but given that the verb root *mih* 'to ask' is a divalent verb without a benefactive/applicative, it must be the latter. Now looking at the thematic + extension form *-taw*, we know that the TAM must be in the remote past. Accordingly, this is an imperfective verb with no dual arguments. The ultimate consonant of the verb form is *-n*, indicating that the actor is first person and singular. This gives the total interpretation of 'I was asking you a long time ago'.

Section 4.7 also gives an exposition of how a circumfixal paradigm works, specifically from the point of view of TAM specification.

Table 4.31 shows elicited combinatorics of u-prefix set with suffixal forms.

U-prefix Set	1sg	1du	1pl	2sg	2du	2pl	3sg	3du	3pl	Unified TAM
β				-ta						Imperfective Imperative
ɑ	-tan	-wm	-tam	-t	-te	-tate	-t	-wt	-tat	Imperfective Non-Prehodiernal
β	-tan	-wm	-tam	-t	-te	-tate	-t	-wt	-tat	Imperfective Yesterday Past
ɑ	-tamn	-mm	-tamm	-tam	-mnde	-tamnde	-tamng	-mnd	-tamnd	Imperfective Primordial
β	-tawn	-wm	-tawm	-taw	-wt	-tawte	-taw	-t	-tawt	Imperfective Remote Past
β				ø						Perfective Imperative
ɑ	-sn	-em	-sm	-sa	-end	-ste	-sa	-end	-st	Perfective Past
	-n	-em	-sm	ø	-end	-ste	ø	-end	-st	Alternative suffix form
β	-sn	-em	-sm	-sa	-end	-ste	-sa	-end	-st	Perfective B
	-n	-em	-sm	ø	-end	-ste	ø	-end	-st	Alternative suffix form
ɑ	-yn	-eaym	-ym	-y	-eaende	-ynde	-y / -yng	-eaend	-ynd	Perfective Preterite
β	-ngn	-em	-ngm	-nga	-end	-nge	-nga	-end	-ng	Perfective Future
β	-ngayn	-eaym	-ngaym	-ngay -ngayng	-eayng -eaynd	-ngaynge -ngaynde	-ngay -ngayng	-eayng -eaynd	-ngayng -ngaynd	Perfective Past Habitual

Table 4.31: Circumfixal verbal paradigm. Combinatorics of u-prefix set type with suffix to provide a unified TAM. The person and number expressed in the row headers (e.g. 1sg, 1du) refer to the actor person and number expressed by the suffix.

4.5.8 Prefixing Verb Particularities

Prefixing and ambifixing verbs share everything left of the verb root in terms of architecture. Prefixing verbs can thus be viewed as reduced forms of ambifixing verbs. The u-prefix has the same two sets of person and number markers as ambifixing verbs (section 4.5.3). Prefixing verbs also make use of the directional prefixes (4.5.4), and the diathetic forms part of the stem (4.5.2). Unlike ambifixing verbs which rely on the thematic + extension to specify various TAM values, prefixing verbs can express only four tense distinctions. The four tense distinctions belong to the imperfective paradigm (4.7).

There are three types of prefxing verbs:

- Verbs that are built up from the copula: 'come', 'go', and 'own'.
- The 'walk' verb with a person/number invariant suffix that looks like the first singular non-dual imperfective actor suffix (*-tan* on ambifixing verbs).
- 'Positional verbs' that mark the sole argument through the u-prefix, and that also take a special set of suffixes that mark dual number. As the name suggests, the verb semantics of positional verbs concern the position the sole argument is taking in relation to its surroundings,

For the remainder of this section I will elaborate on the particularities of these three types.

4.5.8.1 Copula and copula based verbs

The most frequently occurring type of prefixing verb is the copula. The sole argument is indexed on the u-prefix. The verb root suppletes between non-dual $\sqrt{m}$ and dual $\sqrt{ren}$. Copulas are used in presentational, and existential constructions. There is noticeable and consistent variation in the form of the copulas where a final [n] is attached to the non-dual form, e.g. *ym* and *ymn* both 3sg non-prehodiernal (nphd) forms. Table 4.32 shows the copula forms with optional [n] where it has been observed in the corpus. There is no known meaning variation to the [n]-ful forms.

S Argument	Nphd	Ypst	Primordial	Remote Past
1sg	wm(n)	qm(n)	wmaro(n)	qmaro(n)
1du	ynre(n)	tnre(n)	ynrmn	tnrwm
1pl	ynm	tnm	ynmaro(n)	tnmaro(n)
2sg	nm(n)	km(n)	nmaro(n)	knmaro(n)
2du	ere(n)	tare	ermn	tnrwm
2pl	em(n)	tnm	emaro(n)	tnmaro(n)
3sg	ym(n)	tm(n)	ymaro(n)	tmaro(n)
3du	ere(n)	tare	ermn	tarown
3pl	em(n)	tam(n)	emaro(n)	tamaro(n)

Table 4.32: Conjugation of the copula verb.
nphd = non-prehodiernal ; ypst = yesterday past.

Like ambifixing verbs, the copula can be modified with the directional affixes venitive *n-* 'hither' and andative *ng-* 'thither'. The resulting semantics are verbs of motion 'come' and 'go'.

(124) Contrastive examples of directional prefix

a. ynd mer w/m.

1ABS good 1sg.U:nphd/-be

I am good.

b. w-**n**/m

1sg.U:nphd:**ven**/be

I came.

c. w-**ng**/m

1sg.U:nphd:**and**/be

I went.

Copulas can also express large plural number like ambifixing verbs by appropriating the *ng-* prefix. This results in homophony of the 'go' form of a verb and the large plural form of the copula. In most cases the context disambiguates, but sometimes more careful interrogation and translation is required. In the example below it is not entirely clear whether whether *wngm* is the large plural (giving the interpretation 'My lack of knowledge is great'), or whether it is the 'go verb' and some metaphorical expression of knowledge leaving the speaker. In this instance I analyse it as a copula with the large plural, since the expression 'I don't know' is formed by the collocation of *mavña* + COP.

(125) a. [Alqi is talking about her Idi language comprehension.]

Ämb	got	yao,	mavña	gs	**w-ng/m**,
some	word	NEG	lack.knowledge	DEM	1pl+.U:nphd/be

yao	ge	n-ng\äyäre/tan.
NEG	DEM	1sg.A:ipfv.nphd:and/hear

Some words, no, I don't know, I can't hear them.

WSEK1-G20170617-01Alqi02MQ, 00:17:32.174 - 00:17:36.974

Evans (2019c) notes for Nen that the 'come' and 'go' verbs as having large plural suppletive forms. I was only able to elicit a few forms for Nmbo (see below), but it seems this paradigm only utilises the singular forms of the u-prefix. It appears to form a paradigm for all four imperfective TAMs. Note that the final syllable of the elicited remote past form, [nao] resembles the third singular form of an ambifixing verb, showing the appropriate morphophonemics when the verb root ends in [n] (section 4.5.5). In natural speech, however, we find the non-morphophonemicised form *tao* (example 4.5.8.1).

We will see with the 'walk' verb (section 4.5.8.2) also where parts of the verb template show forms that appear intermediary between imperfective ambifixing

verbs and prefixing verbs.

- *w/äwalman* - 'We go'
- *w-n/äwalman* - 'We go'
- *y/äwalman* - 'They go'
- *y-ng/äwalman* - 'They have already gone past'
- *t/äwalman* - 'They went' (Yesterday Past)
- *t/äwalmanao* - 'They went' (Remote Past)

(126) a. Kkv=at yao bä t-ngäwälm**tao**
garden=ALL NEG FUT 3sg.U:rmpst:and/go
They will not go to the garden.
WSEK1-G20170707-03MY02MQhm, 00:10:25.559 - 00:10:26.711

The final copula based verb is what is translated as 'to own' or 'to have'. It is built up from the copula by the diathetic benefactive/applicative *(w)awa-*. The person and number of the possessor are indexed on the u-prefix. The examples I have are all with the possessor expressed in the possessive pronominal form (or the possessor noun with a possessive suffix), and the possessed item is not overtly expressed.

(127) Contrastive examples of directional prefix

a. bende n/awa-m
2sg.POSS 2sg.U/ben.app-be
It's yours.

b. yvende sombwi aï-vende e/wawa-re
3nsg.POSS two person-POSS.nsg 2du.U/be/app-be
It's the two men's / It belongs to the two men.
Fieldnotes 2017 Book 2:7

S	Nphd	Ypst	Primordial	Remote Past
1sg	wawam	qawam	wawamaro	qawamaro
1du	ynaware	tnaware	ynawarwn	tnawarwn
1pl	ynawam	tnawam	ynawamaro	tnawamaro
2sg	nawam	knawam	nawamaro	knawamaro
2du	ewaware	taware	ewawarwn	tawawarwn
2pl	ewawam	tawam	ewawamaro	tawamaro
3sg	yawam	tawam	yawamaro	tawamaro
3du	ewaware	taware	ewawarwn	tawawarwn
3pl	ewawam	tawam	ewawamaro	tawamaro

Table 4.33: Conjugation of the 'own' verb.

4.5.8.2 'Walk' verb

The 'walk' verb has a suppletive root $\sqrt{ta}$ for non-dual and $\sqrt{w}$ for dual, which resembles the alternation forms of the imperfective thematic for non-prehodiernal and yesterday past tenses. Unlike the other prefixing verbs, the semantics of this verb are dynamic rather than stative.

As mentioned above for the suppletive large plural 'come/go', the morphology of 'walk' shows what appears like an intermediary form between ambifixing verbs. The final portion of the verb *-tan* is reminiscent of the first singular actor in non-prehodiernal and yesterday past paradigms for ambifixing verbs. Unlike the ambifixing verbs, however, the 'walk' verb *-tan* shows no sensitivity to the person/number of the sole argument (table 4.34). The thematic *-m* for primordial and *-w* for remote past appear to be functioning, and the u-prefix will take the appropriate set: the α-set for primordial TAM, and the β-set for the remote past.

(128) a. q-u-ta-wn...

1sg.U:β-walk-IPFV.DU-RMPST

I walked...

WSEK1-B20150721$_0$1marriageAT, 10:22.940 - 10:26.732

b. bä t-u-ta-wn...

3.ABS 3sg.U:β-walk-IPFV.NDU-RMPST

She walked...

WSEK1-B20150805-01NinyiTS, 02:20.075 - 02:22.285

c. Sävuka Maraga bä ta-yu-ø-wn Wando=t.

Sävuka Maraga 3ABS 3nsg.U:β-walk-IPFV.DU-RMPST Wando=ALL

Sävuka and Maraga walked to Wando.

WSEK1-B20150721$_0$1marriageAT, 09:46.557 - 09:48.993

S	Nphd	Ypst	Primordial	Remote Past
1sg	wuyutan	quyutan	wuyutamn	quyutawn
1du	yñuwn	tñuwn	yñumn	tñuwn
1pl	yñutan	tñutan	yñutamn	tñutawn
2sg	nuyutan	kuyutan	nuyutamn	kuyutawn
2du	eyuwn	täyuwn	eyumn	täyuwn
2pl	eyutan	täyutan	eyutamn	täyutawn
3sg	yuyutan	tiyutan	yuyutamn	tiyutawn
3du	eyuwn	täyuwn	eyumn	täyuwn
3pl	eyutan	täyutan	eyutamn	täyutawn

Table 4.34: Paradigm for the 'walk' verb. nphd = non-prehodiernal ; ypst = yesterday past.

A final comment on a bit of variation observed for the 'walk' verb. My consultant Bunai Äniba rejected the form *ynd ynyutan* 'we (plural) walked', and suggested the form *ynd ynyuta* without the final [n]. It is not clear whether his objection was to the final [n] in this particular verb context, or with the entire paradigm. It is a curious comment given that other prefixing verbs such as the copula and positionals show variation in realising word final [n]. Various speculations can be made as to whether this judgement was a hyper-correction (i.e. the copula should not have a final [n], therefore any verb of a prefixing paradigm should not have [n]), or if he has some awareness of older forms that were possibly used by older speakers in the past (i.e. the 'walk' verb's final *-tan* is a historical innovation).

4.5.8.3 Positionals

The prefixing verbs which Evans (2015c) calls *positionals* are identifiable by their unique semantics and suffix (p.1079). Positionals are a limited set of verbs which express stative positions such as 'to be sitting', 'to be hanging', 'to be wedged between something'.

The positional suffixes are sensitive to a non-dual/dual distinction. There are two suffixes, *-(o)ngr* for a non-dual, and *-ara* for dual. Singular person/number of the sole argument is expressed by a singular u-prefix and non-dual suffix, and dual number by a non-singular prefix and dual suffix. Plural number is expressed by a non-singular prefix and non-dual suffix. Positionals can express large plural number, but unlike other verbs it does so by combining the singular u-prefix and the dual form of the suffix (see table 4.35).

Positionals, like all prefixing verbs, make four TAM distinctions. The primordial and remote past tenses are formed with the thematic extensions *-m* and *-w* respectively, and takes the appropriate u-prefix ; u-prefix in the α-set for the primordial, and the β-set for the remote past.

S person/number	Nphd	Variation	YPst	Primordial	Remote Past
Singular	wakiongr	wakiongrn	qakiongr	wakiongrmn	qakiongrwn
Dual	ynakiara	ynakiaran	tnakiara	ynakiarmn	tnakiarwn
Plural	ynakiongr	ynakiongrn	tnakiongr	ynakiongrmn	ynakiongrwn
Large Plural	wakiara	wakiaran	qakiara	wakiarmn	qakiarwn

Table 4.35: Positional circumfixal paradigm demonstration with the verb $\sqrt{aki}$ 'to be standing'.

Positional verbs are a limited set. There are 45 positional verbs observed in Nen (Evans pers.comm.), 41 in Komnzo (Döhler 2018:193), and Nmbo is likely to have a similar number. Positional verb roots are unique to this class and cannot be the root of any other form, e.g. *yvyongr* 'it is hanging' **fyoh*. There are, however, some exceptions such as ykmongr 'it is laid down' *km-h* 'to lay down; sleep', *ydñingr* 'it is hidden', dñi-h 'to hide something'. A list of 24 known positional verbs is presented in the appendix (A.2).

(129)

a. [Positional inflection of *kmh* 'to lay down; sleep']

k-n\wan/em	e	kt	marwa	bu	**y-kmo/ngr**.
1du.U:pfv.proxin/depart	DISC	there	tree.type	log	3sg.U-lay.down-ndu.nphd

The two of us departed to where the marwa log is lying.

WSEK1-B20150817-02DimbanKSae, 00:06:43.165 - 00:06:47.336

b. [*kmh* in an ambifixing construction]

banban	mna=yan	trei=an	**y\kmo/yng**.
below	lie.so=LOC	tray=LOC	3sg.A>3sg.U:pfv.pret/lay.down

He laid it down like so at the bottom of the tray.

WSEK1-B20150909-MissingMoneyMZ, 00:31:24.717 - 00:31:26.557

Positionals, like the copula, are sometimes realised with a final [n]. Speakers do not ascribe any difference in meaning to these form, so is treated as an alternative variant to the standard positional suffix.

(130) a.

wagiv	smi-e	ge	y-kmo-ngr**n**.
fish	tail-DAT	DEM	3sg.U-lay.down-ndu.nphd

The fish tail is lying here.

WSEK1-B20150817-02DimbanKSae, 00:05:31.230 - 00:05:33.383

4.6 Noun Phrase Structure

The Nmbo noun phrase (NP) is a group of nominals which collectively fulfil a particular function in the clause. They may function as arguments or adjuncts of a predicate.

The strictness of internal constituent order will not be considered in much depth, aside from a few comments here. The current analysis is that NP constituents (or slots) tend to occur in certain orders, following the same template presented in Döhler (2018) for Komnzo. There are, however, plenty of exceptions within

the corpus of slots occurring in different orders. For the time being I will present a template displaying common slot orders as an abstraction, with the understanding that there is variation to the order. Examples of variable slot order will be visible in the examples throughout this section.

Each NP has a single head, which can be a common noun, a personal pronoun, an interrogative pronoun, a compound word, or an infinitive verb. At a minimum, the NP may be the head by itself. This head can, however, be omitted. Nmbo clauses are therefore often a single inflected verb, where pragmatics or indexation on the verb provides information about the arguments.

There are three slots + the head slot for a total of four slots, and case marking occurs at the right edge of the NP. All slots are optional. As Döhler notes of Komnzo (2018:266), NPs rarely have all slots realised in a single NP. The preference in natural speech is to distribute information across the clause, as independent NPs preceding and following the VP (Döhler 2018: 271).

Determiner (DET)	Modifier 1 (MOD1)	**Head**	Modifer 2 (MOD2)
Adnominal DEMs	Numerals	Personal Pronouns	Spatial DEMs
Indefinite Pronouns	Quantifiers	Indefinite Pronouns	
Possessive Pronouns	Attribute Nouns	Common Nouns	
NPs of Character*		Compound Words	
NP of Possession		Pronominal DEMs	
		Nominalised Verb	

Table 4.36: Template for Nmbo NP.
NPs that are modified by the semantic cases:
ORIginative, PRIVative, COMitative

Case markers occur at the left edge of the NP. Examples 131a and b were deemed semantically equivalent and grammatically acceptable by two consultants, yielding the interpretation 'the small dog bit the large pig'. In example b the interpretation is still that the dog is the actor, not the 'small-ness'.

(131) a.

[nmbnarvi	aha-m]$_{NP}$	[mamwi	kitong]	trn.
small	dog-ERG	pig	big	3sg.A>3sg.U:pfv.proxin/bite
MOD1	head	head	MOD1	VP

The small dog bit the large pig.

b. [aha nmbnarvi-am]$_{NP}$ [mamwi kitong] trn.

[dog small]$_{NP}$-ERG [pig big] 3sg.A>3sg.U:pfv.proxin/bite

head MOD1 head MOD1 VP

The small dog bit the large pig.

Fieldnotes 2017: 68

The *determiner* (DET) slot can be occupied by demonstratives functioning adnominally (section 4.2.1.3), indefinite pronouns, possessive pronouns (4.2.1.1), and NPs with specific case markers. Example 132 shows *yna* DEM.prox preceding the head noun *ag* 'coconut', and the indefinite pronoun *na* 'what' preceding *dvn* 'root', with the locative case marking the right edge of that NP.

(132) a. [bmo] [**yna** ag] ge i\ye/i,

[2sg.ERG]$_{NP}$ [**DEM.prox** coconut]$_{NP}$ DEM 2sg.A>3sg.U:pfv.pret/plant

[**na** dvn=an] i\ye/i?

[**how** root=LOC]$_{N}$P 2sg.A>3sg.U:pfv.pret/plant

You planted this coconut, why did you plant it?

WSEK1-G20151015-06BurnedMN, 00:00:14.035 - 00:00:16.443

b. [**nmbone** nne] yna ta\neh/awt

[**small** food]$_{NP}$ DEM.prox 3nsg.A>3nsg.U:ipfv.rmpst/cook

dev=av.

father=DAT.nsg

A small feast was cooked up by the fathers.

WSEK1-B20150721$_0$1marriageAT, 00:13:20.916 - 00:13:23.221

NPs with specific case markers can occupy the DET slot. Some of the most common are NPs marked with an possessive (section 4.4.4.3), or an originative (4.4.4.15). These NPs of character often have a single slot occupied, as in example 133a, but can also be an NP with more than one slot expressed as in examples (b) and (c).

(133) a. [tande ama] gym, [[Govav=mn] mrz] y/m.
1sg.POSS mother TOPIC Govav=ORI daughter 3sg.U:nphd/be
DET head

My mother [TOPIC], is a daughter of Govav.

WSEK1-G20170713-02MQMW, 00:01:19.648 - 00:01:21.591

b. bm [[bende ama=nde] zi] ge n\owav/t...
2ABS 2sg.POSS mother=POSS language DEM 2sg.U:ipfv.nphd/speak
[[DET head=POSS] head]

You speak your mother's language.

WSEK1-G20170617-01Alqi02MQ, 00:15:31.626 - 00:15:33.763

c. [[[ama=nde] hamba=mn] är] ge y/mn.
mother=POSS village=ORI person DEM 3sg.U:nphd/be
[DET]head]

Mother's people, they are coming.

WSEK1-G20170617-01Alqi02MQ, 00:08:31.660 - 00:08:33.661

The *Modifier 1* (MOD1) slot can be occupied by numerals (example 134a), quantifiers (134b), adjectives and common nouns being used attributively (134c). Examples (a) and (c) show typical order with the MOD1 slot occurring to the left of the head, but example (b) shows an atypical order with the quantifier following the head.

(134) a. [sombwi svälnḡ] ere.
two bag.ABS 3du.U:nphd/be
MOD1 head

There are two bags.

WESK1-B20170714-02Coconuts04SZ, 00:04:27.226 - 00:04:28.155

b. [är ämb] ge=ymn... yukr e/m.
people.ABS some TOPIC absent 3nsg.U:nphd/be
head MOD1

Some people [TOPIC] are missing

WSEK1-A20150917-1SinkingBA, 00:26:17.053 - 00:26:19.080

c. Ag gym, [kitong rokar] y/m.
coconut.ABS TOPIC large thing 3sg.U:nphd/be
MOD1 head

A coconut [TOPIC] it is an important thing.

WSEK1-G20141022-01CocoGMh88, 00:05:30.004 - 00:05:31.707

The *head* is both the syntactic and semantic head of the NP. Semantically it provides information about the whole phrase, and syntactically it is indexed on the main verb of the clause. The head lot can be occupied by personal, interrogative, and indefinite pronouns (sections 4.2.1.1, 4.2.1.6, 4.2.1.7), common nouns (section 4.2.1.2), compound words (example 135a), infinitive verbs (section 4.4.1, example 135b), and demonstratives used pronominally or anaphorically (example 135c).

(135) a. [mora-mora mngo=t] wa e-ng/m...
medicine.ABS house=ALL must 3pl.U:nphd/be
MOD1 head

You must go to the hospital.

WSEK1-G20150826-03AusWorkZG, 00:06:54.930 - 00:06:56.891

b. [äuw-h kitong-am] y\räme/t yädm.
[happy-INF large-ERG] 3sg.A>3sg.U:ipfv.nphd/do 3sg.POSS.mother
head MOD1

[lit. Great happiness was affecting her mother.]

WSEK1-A20170616-01RM04Retellhm, 00:03:49.613 - 00:03:51.511

c. är ynandma, [ämb tvende] e-ng/maro.
people.ABS DEM.prox.ABL some 1nsg.POSS 3pl.U:nphd/go
head DET

People went there, some of ours went.

WSEK1-A20150923-07TzZiBA, 00:02:38.116 - 00:02:39.758

The head of a NP can be elided. Example 136 shows the elision of the head, and only the modifer *tärvär* 'plenty' is expressed. Example 137 shows a typical example of a full NP elision, where one of the core arguments of the predicate is elided (in this case, the agent argument).

(136) [Elided head example: the elided head is shows in parenthesis ().]

a. [Wäwi is talking about how she used to hunt big animals when she was a young woman.]

yndo [tärvär (wrar)] t-ng\ñ/awn,
1sg.ERG plenty (animals.ABS) 1sg.A>3pl+.U:ipfv.rmpst:and/kill
MOD1 (head)
äuyä...
cassowary.ABS

I killed many of them, like cassowaries...

WSEK1-B20150924-HuntingWZ, 00:02:39.118 - 00:02:41.689

(137) [Elided NP example: the elided NP is shows in parenthesis ().]

a. [(Yndo)] ḡḡn t\akavo/n.
[(1sg.ERG)] bow 1sg.A>3sg.U:pfv.proxin/pick.up
(head)

(I) got my bow and arrow.

WSEK1-G20150902-03PigGS, 00:01:52.700 - 00:01:54.370

Finally, *modifier 2* (MOD2) is reserved for demonstratives of space/location. This slot is posited because locational demonstratives are not modifying the head adnominally, nor are they functioning pronominally as the head. Instead these locational demonstratives are pronominally standing in for locations that are providing optional information about the NP and its actions.

(138) a. Mato [rokar ynama] bä k\ogamne/nga
perhaps [thing DEM.prox.ABL] FUT 3sg.A:pfv.fut/come.across
head MOD2
de, mamwi o rusa.
ALR pig CONJ deer

Maybe he will meet something along the way, like a pig or deer.

WSEK1-G20150902-03PigGS, 00:02:27.020 - 00:02:30.710

b. [är ynandma], ämb tvende e-ng/maro.
people DEM.prox.ABL QNT 1nsg.POSS 3pl.U:nphd/
head MOD2

People went there, some of ours went.

WSEK1-A20150923-07TzZiBA, 00:02:38.116 - 00:02:39.758

4.7 Inflectional Tense, Aspect, and Mood

This section is an overview of verbal inflection. A simplified explanation of verbal affix combinatorics will show how TAM is specified in a circumfixal paradigm, and clear cases of TAM category semantics will be discussed. The labels for these TAM categories, and the sections in the thesis covering them, are listed in table 4.37.

Aspect	TAM Category	Abbreviation	Thesis Section
Imperfective	Imperative	IMP	4.7.0.1
Imperfective	Non-Prehodiernal	NPHD	4.7.0.2
Imperfective	Yesterday Past	YPST	4.7.0.2
Imperfective	Primordial	PRIM	4.7.0.3
Imperfective	Remote Past	RMPST	4.7.0.3
Perfective	Imperative	IMP	4.7.0.1
Perfective	Past	PST	4.7.0.4
Perfective	Proximate Inceptive	PROXIN	4.7.0.4
Perfective	Preterite	PRET	4.7.0.5
Perfective	Future	FUT	4.7.0.6
Perfective	Past Habitual	PHAB	4.7.0.6

Table 4.37: List of TAM categories and their abbreviations, and relevant sections in the thesis.

The TAM paradigms are fundamentally built upon a distinction of verbs as *imperfective* or *perfective* in aspect. These two terms are used in a slightly Yam language-specific way, but fundamentally follows the classification of imperfectives as paying attention to internal temporal structures (Comrie 1976:23-4), and perfectives as viewing a situation as a whole with bounded temporality (Comrie 1976:17, Bybee, Perkins, and Pagliuca 1994:54). In Nambu languages more specifically, imperfective verbs are characterised as durative, and perfective verbs are momentaneous and punctual (Siegel 2014b). In addition, both aspects focus on the inception of an event rather than its completion (Evans, Arka, et al. 2018:689, Siegel 2014b:211 for Nama).

Many verbs in Nmbo can form both imperfective and perfective paradigms, but there are also many verbs which can only form one. For example prefixing verbs can only form an imperfective TAM paradigm. For non-prefixing verbs, aspect seems to be specified lexically. For example *tereh* 'to hold something'

can be both aspects, but *arengh* 'to wander' only seems to form an imperfective paradigm. The verb *áwih* 'to fall' only seems to form a perfective paradigm. There is a morphological strategy, however, for coercing verbs into an imperfective interpretation, and this is covered in the relevant sections (sections 4.5.5, 4.7.0.6).

The following examples and their translations give a sense of the imperfective and perfective aspectual distinction.

(139) a. [Imperfective]

n\armbo/tan.

1sg:**ipfv**.nphd/ascend

I am ascending [e.g. multiple steps, like up a ladder].

Field notes 2017, Book 1:45

b. [Perfective]

n\armbo/n.

1sg:**pfv**.pst/ascend

I ascended [e.g. a single step, like onto a chair].

Field notes 2017, Book 1:45

As discussed in the ambifixing verb section (4.5.7), the full set of argument and TAM specifications can only be understood when all morphemes of the verb complex are viewed as a whole. Here I will briefly explain how TAM is specified by going through discrete stages of combining the parts.

Firstly, the u-prefix set (4.5.3) and the verbal aspect are combined to limit TAM possibilities. The forms of the u-prefixes according to set is shown in table 4.24 in section 4.5.3. Figure 4.17 schematises how each prefix set and aspect combination yields a minimum of two TAM possibilities. For example, if the u-prefix is part of the ɑ-set, and the aspect is imperfective, the verbal TAM is either the non-prehodiernal (NPHD), or primordial (PRIM).

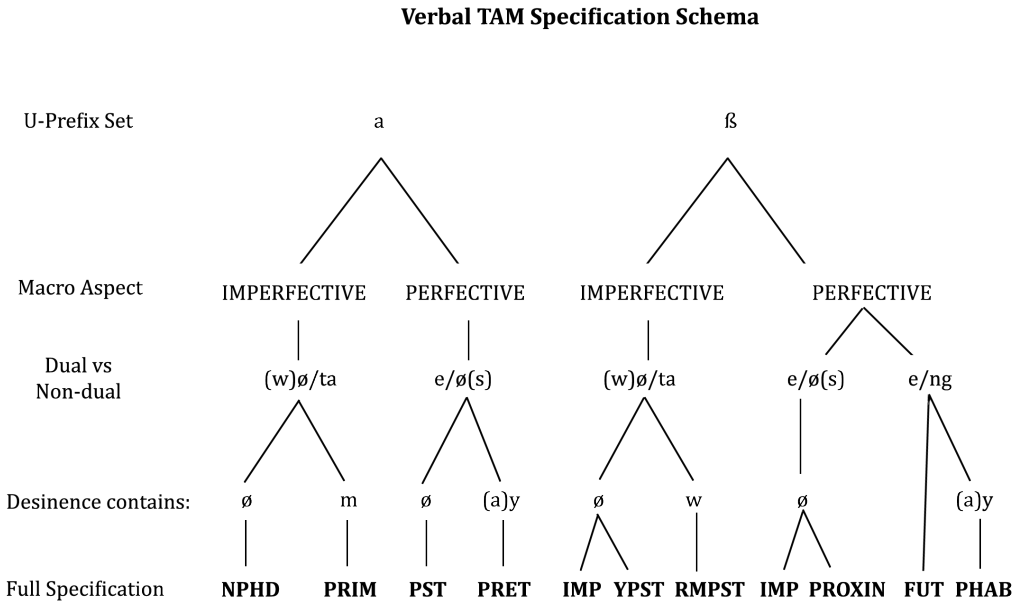

Figure 4.17: Verbal TAM specification schema. NPHD = non-prehodiernal; YPST = yesterday past; PRIM = primordial past; RMPST = remote past; PST = perfective past; PROXIN = proximal inceptive; FUT = future; PHAB = past habitual

Once the possible TAM range is limited, it is further specified by the form of the desinence (i.e. the combined suffixes right of the verb root). For example, the ɑ-set with a perfective aspect combined with a desinence containing *-(a)y* will be a preterite. The β-set with a perfective aspect combined with a desinence containing *-ng* will be the future, or past habitual. The form of the a-suffixes will finalise the verbal TAM. For ease of reference, table 4.38 is a copy of a table from section 4.5.6 presenting forms of the desinence.

4.7.0.1 Imperatives: Imperfective and Perfective

Basic imperatives in Nmbo can occur in both imperfective and perfective paradigms, and always occurs with the β-set of u-prefixes. The imperfective form commands continuation of an action, while the perfective form commands what can be construed as a single action. Imperatives inflect for subject number, which is indexed on the a-suffix. For transitive imperatives, the second argument is indexed on the u-prefix (table 4.39). Nen has an attested future imperative, but this has not yet been investigated for Nmbo.

	Imperfective			Perfective			
	NPHD, YPST	PRIM	RMPST	PST, PROXIN	PRET	FUT	PHAB
1sg	-tan	-tamn	-tawn	n /-sn	-yn	-ngn	-ngayn
1du	-wm	-mm	-wm	-em	-eaym	-em	-eaym
1pl	-tam	-tamm	-tawm	-sm	-ym	-ng-m	-ngaym
2sg	-t	-tam	-taw	-sa /ø	-y	-nga	-ngay -ngayng
2du	-te	-mnde	-wt	-end	-ea-ende	-end	-eaynd -eayng
2pl	-tate	-tamnde	-tawte	-s-te	-ynde	-nge	-ngaynde -eaynge
3sg	-t	-tam -tamng	-taw	-sa	-y -yng	-nga	-ngay -ngayng
3du	-wt	-mnd	-t	-end	-eaend	-end	-eaynd -eayng
3pl	-tat	-tamnd	-tawt	-st	-ynd	-ng	-ngaynd -ngayng

Table 4.38: Copy of table 4.30. Observed desinence forms.

	Intransitive		Transitive	
addressee number	Perfective	Imperfective	Perfective	Imperfective
SG	kitro "You move!"	kitrota "You keep moving!"	tere "You hold it!"	tereta "You keep holding it!"
DU	kitrend	kitronge	terend	ternge
PL	kitronge	kitrotange	terenge	teretange

Table 4.39: Examples of imperatives. For transitive imperatives, it is assumed that the u-prefix will inflect in number for the second argument following the β-set paradigm.

4.7.0.2 Imperfective Non-Prehodiernal and Yesterday Past

The *non-prehodiernal* (NPHD) and *yesterday past* (YPST) TAMs are formed with the same desinence patterns, but with alternate u-prefix sets. NPHD is formed with the ɑ-set of prefixes, while YPST is formed with the β-set.

Following Evans (2015c) for Nen, "non-prehodiernal" refers to any time from this morning on into the future. It is used for present situation, and also used for habitual expressions (example 140). When the future particle *bä* precedes a verb inflected for NPHD TAM, the meaning is future (140c).

Yesterday past, as the name suggests, refers to events that happened a day or so prior. It seems likely that YPST is a relative past tense of sorts, where the time frame is not restricted to the day prior. While in the field one afternoon, the villagers of Bevdvn received news that a woman had passed away suddenly that morning. A young girl from the village said to me *tande mwidädem tm*, "She was my auntie", where the 3sg copula is in the YPST form. If verbal inflections were absolute tenses, the copula would have been in the NPHD form *ym*, but this was not the case. The following example 150, however, shows clear uses of the YPST referring to events that happened one day prior to the utterance.

(140) [Non-Prehodiernal]

a. [Towal is describing what she did earlier in the day prior to this recording. A large portion of the speech is in the perfective past tense, but there are a few examples of the NPHD.]

dena	"**yn-ng/m**	Nmbo	Zi	mngo=t."	Yndon
QUOT	1pl:**nphd**/go	Nmbo	Language	house=ALL	1sg.ERG
bva	**e**\w-laowe/**tan**			rokar-rokar	mngo
just	1sg.A>3pl.U:ipfv.nphd/APP-put.in			thing-RDP	house
knd-n					
inside=LOC					

I said, "We go to the Nmbo Language House." I put the things [cooking utensils] inside the house."

WSEK1-A2015100-12ThisMorningTT, 00:01:59.875 - 00:02:05.967

b. [Habitual use.]

ynd	dumen	kkv=at	**n**\oreñ/**tan**.
1ABS	HAB	garden=ALL	1sg.A:**ipfv.nphd**/clear

I always clear the garden.

Nmbo Dictionary Draft Version 2018, entry *dumen*

c. [Modified with the future particle for future TAM]

kai	kkv	bä	**y**\rte/**tam**.
± 1day	garden	FUT	1sg.A>3sg.U:**ipfv.nphd**/clear

Tomorrow we will clear the garden.

Nmbo Dictionary Draft Version 2018, entry *rth*

(141) [Yesterday Past]

[Rachel is speaking about the events of yesterday as part of the sociolinguistic interview]

a.

ynd	**q**-ng/m	yäme	**t**\anrs/**an**,	karv
1ABS	1sg.U:**ypst**/go	mat	1sg.A>3sg.U:**ipfv.ypst**/bring	oven

t\nend/**an**,	**ta**\ne/**tam**...
1sg.A>3sg.U:**ipfv.ypst**/open	1sg.A>3pl.U:**ipfv.ypst**/eat

I went, I brought the mat, I opened the oven, we were eating the foods...

WSEK1-A20160722-04RM01Ypsthm, 00:03:46.500 - 00:03:49.422

4.7.0.3 Imperfective Primordial and Remote Past

The *primordial* (PRIM) and *remote past* (RMPST) TAMs indicate that an event happened long ago in the past. PRIM occurs with the ɑ-set of prefixes, while RMPST occurs with the β-set. The thematic extension is *-m* for PRIM, and *-w* for RMPST, but the a-suffix form for the 2nd and 3rd persons differ. Both TAMs are used in reference to events from a year ago, right through to the ancestral past. For Nen, PRIM is described as indicating events which occurred

for the first time (Evans 2015a:558), and this seems to be true of Nmbo (example 142).

Pers/Num	Primordial	Remote Past
1sg	-tamn	-tawn
1du	-mm	-wm
1pl	-tamm	-tawm
2sg	-tam	-taw
2du	-mnde	-wt
2pl	-tam(n)de	-tawte
3sg	-tamng	-taw
3du	-mnd	-t
3pl	-tamnd	-tawt

Table 4.40: Desinence forms for the imperfective primordial, and remote past.

(142) [Primordial and Remote Past examples from natural speech.]

[Gvai is explaining how he decided to go hunting. Some segments of speech are abridged for brevity, to show how Gvai decided to go hunting for the first time that day.]

a. zite-n bä **k**\end/**ao**... mna=t,
afternoon=LOC FUT 3sg.A:**ipfv.rmpst**/cry DEM.fw=ALL
wrar mende **t**/**maro**. Okei ynd qévt-an
meat desire 3sg.U:**rmpst**/be DISC 1ABS night=LOC
n\emn/**amn**... modwa=wt ä n-ng\re/tan...
1sg.A:**ipfv.prim**/think bandicoot=PURP FUT 1sg:ipfv.nphd/get

In the afternoon he was crying... because he wanted meat. So that night I thought for the first time... I will go hunt for bandicoot.

WSEK1-G20150902-03PigGS, 00:01:31.941 - 00:01:49.250

4.7.0.4 Perfective Past, and Proximate Inceptive

The perfective past (PST) and perfective proximate inceptive (PROXIN) share desinence forms, but PST is formed by ɑ-set suffixes while the PROXIN is by

the β-set. Both show a dual/non-dual alternation of *-e/-s*. The non-dual *-s* has an alternative zero form for singular number actors across all three persons. In the 2-3sg actor forms there is a *-sa* vs. zero alternation, while for the first person singular and dual forms there is a [s]-ful form vs a not [s]-ful form, e.g. *-sn* vs *-n*. There is no such variation in the second and third plural forms (table 4.41).

Actor Pers/Num	Desinence Form
1sg	-sn ~-n
1du	-em
1pl	-sm ~-m
2|3sg	-sa ~ø
2|3du	-end
2pl	-ste
3pl	-st

Table 4.41: Variant desinence forms for the perfective past and perfective proximate inceptive.

The perfective past indicates a punctual event that has just occurred. It is also used to describe events that happened earlier in the morning, showing some temporal overlap with the imperfective NPHD (example 143a.). The perfective past is formed with the ɑ-set of suffixes.

(143) [Perfective Past]

[Towal is describing what she did earlier in the day prior to this recording. Note she uses both the *s* and non-*s* suffix variants of the PFV.PST]

a. (b)va yn-n\m... mgo=t **n**\ovar/**sm**... svälnḡ
just 1nsg:nphd/come house=ALL 1sg:**pfv.pst**/arrive yam.basket
y\msu/**n**...
1sg.A>3sg.U:**pfv.pst**/put.down

We just came... We arrived at the house... I put down the basket.

WSEK1-A2015100-12ThisMorningTT, 00:00:30.607 - - 00:00:36.115

The label for the Nmbo proximate inceptive TAM category comes from Nama, which has very similar semantics: "Proximate inceptive tense focuses on the starting point of activities or events that occurred in the not-so-distant past, especially in narratives, and also those that will begin some time in the future." (Siegel 2014b:223). Since the Nmbo corpus mostly constitutes narratives, most examples of PROXIN available are in the narrative past use.

The paradigm of the Nmbo proximate inceptive is somewhat similar to the Nama one barring the forms of the desinence which are rather different.

(144) [Proximate Inceptive: Narrative past use, or starting point of activities or events that occurred in the not-so-distant past]

a. tn-n/maro e~ä n\ovngo/ym urävär- h
1p.Ul:rmpst/come DISC FUT 1pl:pfv.pret/do short.cut-INF
bäi. Eric yndo tnd-e **t**\ramo/**n**
DISC Eric 1sg.A hand-DAT 1sg.A>3sg.U:**pfv.proxin**/make
dena de,
DEM.mn ALR

We came~and we started to make a short cut. I started gesturing to Eric like this [lit. With my hand I started making like this to Eric].

WSEK1-A20150917-1SinkingBA, 00:02:24.130 - 00:02:28.215

b. ynd foa-e n-ng\ovar/yn. Kt-anmae
1ABS later-DAT 1sg:pfv.pret:and/arrive DEM.dist-DEM.ABL
k-n\ago/**n**, daukr-e k-n\ango/tawn
1sg.A:**pfv.proxin**:ven/arrive old.garden-DAT 1sg:ipfv.rmpst:ven/return
mngo...
house

I arrived after that. From there I started to return, from the old garden I returned to the house...

WSEK1-G20150902-03PigGS, 00:02:05.790 - 00:02:12.120

(145) [Perfective Proximate Inceptive: Something that will begin in the future.]

a. [Mawava starts telling a story about past customs by stating that she will be beginning.]

yndo	zi.	bä	**t**\rñi/**n**.
1sg.ERG	story	FUT	1sg.A>3sg.U:pfv.proxin:begin

My story. I'm about to start.

WSEK1-B20150727-01PastCustomMZ , 00:00:02.530 - 00:00:04.943

4.7.0.5 Perfective Preterite

The *perfective preterite* (PRET) indicates a punctual event that happened in the distant past. The u-prefix is in the ɑ-set. The thematic extension is *-(a)y*, resulting in a dual/non-dual distinction of *-eay/-(a)y*.

(146) [Elicited examples of the perfective preterite]

a. **y**\avr/**eaym**

1du.A>3sg.U:pfv.pret/do

The two of us did it.

Fieldnotes 2017, Book 1:21

b. **y**\avro/**ym**

1pl.A>3sg.U:pfv.pret/do

We all did it.

Fieldnotes 2017, Book 1:21

(147) [Example from natural speech.]

a. ynd toge däv yna taim-an q\mraon,
1ABS child when DEM.prox time=LOC 1sg:rmpst/be
foyotaro, skool mak=an ge **n**\auwere/**yn**...
fine school centre=LOC DEM 1sg:**pfv.pret**/attend
At the time when I was a child, it was good, I attended school.

WSEK1-G20150831-05WorkingMG, 00:00:56.787 - 00:01:01.409

4.7.0.6 Perfective Future, and Past Habitual

The *perfective future* (PFV.FUT) is formed by the β-set of prefixes. The dual/non-dual distinction of the thematic takes the form *-e/-ng*. PFV.FUT can be preceded by a future particle *bä*, but this is optional (contrast examples 148a and b).

(148) a. yndo ä **t**\nfo/**ngn**...
1sg.ERG FUT 1sg.A>3sg.U:**pfv.fut**/cut
I will cut it.

WSEK1-A20170619-01Coconuts$_0$2TPhm, 00:03:52.800 -00:03:54.400

b. Krismis ge ym, gmb **t**\nfo/**ngn**...
Christmas DEM 3sg.U:nphd/be coconut.shoot 1sg.A>3sg.U:**pfv.fut**/cut
At Christmas, ok I will cut the coconut shoot.

WSEK-1A20170619-01Coconuts$_0$2TPhm, 00:04:28.000 -00:04:30.100

What I am tentatively calling the *perfective past habitual* (PHAB) is built up from the perfective future. It takes the β-set of u-prefixes and makes a dual/non-dual distinction of *-eay/-ngay*. In other words the PHAB looks as though it is formed by the thematic extension *-(a)y* (section 4.5.5) on the perfective preterite (section 4.7.0.5). The semantics also seem compositional. The future *-ng* and past *-ay* combined produces a past anticipatory interpretation, translated in English as 'would'.

Actor Pers/Num	Desinence Form
1sg	e-ay-n
1du	ng-ay-m
1sg	e-ay-m

Table 4.42: Desinence forms for the perfective past habitual (PHAB). The a-suffix is in first person form

(149) [Perfective Past Habitual in natural speech]

a. [Melsy is explaining how she learnt to weave by copying her grandmother.]

yndo	ge	t-ng\ake/tawn,	ynd	woi
1sg.ABS	DEM	1sg.A>3pl+.U:ipfv.rmpst/watch	1ABS	again

kt	ädi	k\ovnmo/**ngayn**...
DEM.dist	EMPH	1sg.A:pfv.phab/copy

I watched her many times, I **would** copy from there.

WSEK1-G20151015-15BasketMW, 00:01:22.796 - 00:01:26.798

b. [Teräv is recounting the time he was delirious from a snake bite, and how his wife spoke to him then.]

tnd=an	**q**-n\ere/**ngayng**.	**q**\amndo/**ngayng**
hand=LOC	1sg.3>1sg.U:pfv.phab:ven/hold	1sg.3>1sg.U:pfv.phab/say

"end	ä	y/m."
road	3ABS	3sg:nphd/be

She would hold me by the hand. She would say to me "there's a road."

WSEK1-B20150804-01SnakeBiteTY, 00:08:32.150 - 00:08:35.800

Verbs inflected for PFV.FUT and PFV.PHAB often have a dual number invariant *-ta* preceding the desinence. These are instances of the *secondary thematic* (section 4.5.5), where perfective verb roots are coerced into imperfective interpretations by the addition of a *-ta*. The form of this extension *-ta* is unaffected by

dual number, so the dual/non-dual alternations follow the respective paradigms of the FUT and PHAB. The translations offered for these verb forms are very similar to those given for PFV.PHAB predicates.

(150) [Perfective Past]

[Mada is talking about a time he was severely injured.]

a. är t-ngäw-end/awng tande=mn ye. Ynd
people 3pl+.A:ipfv.rmpst/RR-cry 1sg.POSS=ORI cry 1ABS
kowavta-ngayn... mé äksn-e=ro.
1sg.A:pfv.phab+ta/speak CONT action=DAT-RST

Many people cried because of me. I would talk only with actions/gestures.

WSEK1-B20150909-MissingMoneyMZ, 00:26:06.243 - 00:26:11.752

4.8 Simple Clauses and Valency

This section concerns the syntax of Nmbo simple clauses. The basic constituent order is SOV (or AUV, actor-undergoer-verb), and overt pronouns are optional. Since argument structure is encoded largely by case marking (section 4.4.4) and verbal morphology (4.5), this section stands as a brief summary of clausal types explored earlier.

4.8.1 Monovalent Clauses

Intransitive clauses can be formed with two types of verbal inflectional morphology. One is by prefixing verbs verbs:prefixingp, and the other through amibifixing verbs with a middle prefix (section 4.5, 4.5.3). The single argument is always in the absolutive case.

(151) a. [Intransitive clause with prefixing verb. The sole argument is cross-references on the prefix.]

ynd **w-**n-m

1ABS **1SG**-VEN-be

I'm coming.

(152) a. [Intransitive clause with middle marked (M) ambifixing verb. The sole argument is cross-references on the a-suffix.]

ynd n-armbo-ta-**n**

1ABS M:ɑ-ascend-pfv:ndu-**1sg**(nphd)

I'm going up.

4.8.2 Divalent Clauses

Transitive clauses are formed with ambifixing verbs as the predicate, which cross-reference both the actor and undergoer arguments. The actor is in the ergative, and the undergoer in the absolutive. The ambifixing verb template is described in section 4.5.1. An example of a transitive clause is given below.

(153) Transitive clause with ambifixing verb.

a. yndo bä **y**-nfwe-ta-**n**

1sg.ERG 3ABS **3sg.U**-cut-NDU:IPFV-**1sg.A**

I'm cutting it.

4.8.2.1 Experiencer Object Constructions

Experiencer object constructions are a subtype of divalent constructions involving expressions of bodily, mental, and emotional processes such as hunger, fear, and cold (the term from Evans 2015c:1083 for Nen) . One of the first descriptions of this construction type in the Papuan language of Kalam notes how these "bodily and mental process" are grammatically expressed as agentive entities, while the person experiencing them are semantically patientive (Pawley, Gi, Majnep, and Kias 2000:153). It is a common type of construction found in other languages of

New Guinea such as Oksapmin (Trans-New Guinea, Loughnane 2009:328), and Mian (Ok family, Fedden 2011:266). The stimulus is ergative-marked and cross-referenced on the a-suffix, and the experiencer is in the absolutive and cross-referenced on the u-prefix. The verbs that form the predicates of experiencer object constructions are drawn from a limited set, including *ramah* 'to do; make', and *wivoh* 'to finish up something'.

(154) a. ynd bérbér-am q\rame/tawt
1ABS fear-ERG 3pl.A>1sg.U:ipfv.rmpst/do
I was scared. [lit. Fear was doing me.]
WSEK1-G20140922-13MarriageGM, 00:01:55.931 - 00:01:58.653

b. ynd wrar kmhava-m w\räme/t,
1ABS meat protein-ERG 3sg.A>1sg.U:ipfv.nphd
I am hungry for meat. [lit. Meat-flavour is doing me.]
WSEK1-B20150924-HuntingWZ, 00:03:47.798 - 00:03:51.420

4.8.3 Trivalent Clauses (Three-place Predicates)

Nmbo three-place predicate clauses index arguments in two ways. One, the presence of the recipient argument (R) can be indicated on the benefactive/applicative prefix immediately left of the verb root. In such cases the person/number of the object being transferred (T) is indexed on the u-prefix (example 155a). The second way is to cross-reference R on the u-prefix, showing agreement of person/number. In this case the existence of T is indicated by the benefactive/applicative (example 155b). Note that in example 155a the dual number of T is not indexed on the verb in any way, but in (b) the benefactive/applicative prefix indicates the presence of T, while the diathetic to the right of the root is in the dual form *-we*.

(155) a.

ymo	sombwi	nne	w\aram/.
3sg.ERG	two	yam.ABS	3sg.A>1sg.U:pfv.pst/give
A		T	

He gave me two yams.

Fieldnotes 2014 *The Book of Collections*: 7)

b.

yndo	Eri**-e**	sombwi	wagiv	y-**a-**ram/we-n.
1sg.ERG	name**-DAT**	two	fish	1sg.A>3sg.U+DU:ipfv.nphd/APP-give
A	R		T	

I gave Eri two fish.

Fieldnotes 2014 *The Book of Collections* :26)

As is visible from the above examples, the case marking of the third argument also varies. In example 156a the free pronoun is a first singular dative *ta* in a recipient role, and is indexed on the undergoer prefix (u-prefix). The theme, *ynḡ* 'bag', is indicated as a third argument through the valency increasing *a*-diathetic prefix immediately to the right of the u-prefix (with the APPlicative label in example 156a). In example 156b the free pronoun is still the first person singular in the role of R, but in the goal form *ta=vav*. Now R is not indexed on the verb, but T *ynḡ* is. The theme is indexed on the verb by the u-prefix, and the free noun is still in the absolutive form. There are, in other words, two argument frames for three-place predicates: one with an applicative, and one without. The frame with the applicative has an indexed object, while the non-applicative frame indexes the indirect object. A similar pattern has been described for another Nambu branch language, Nen (Evans 2015c:1088).

(156) The speaker is talking about the time they were travelling to Daru by dingy.

a.

Ymo	ämbru	ynḡ	w-a-ramo-i	**ta**.
3sg.ERG	one	bag.ABS	1sg.U-APP-give-1sg.A:pfv.pret	**1sg.DAT**

He gave me one bag.

b. Ymo ämbru ynḡ y-ramo-i **ta-vav**.
3sg.ERG one bag.ABS 3sg.U-give-3sg.A:pfv.pret **1sg.DAT=GOAL**
He gave me one bag.

Fieldnotes 2017 Book2, p.9

4.8.4 Valency Increasing Constructions

4.8.4.1 Causative Constructions

Causative constructions in Nmbo introduce a causer argument to an otherwise monovalent occurrence by adding the diathetic prefixed *w-* on the root. This causative prefix makes the S argument a P argument, and the causer takes the position of the A argument. The nominal referent of the causer takes ergative case, and is indexed on the actor-marking a-suffix. The S turned P argument is indexed on the u-prefix. Example 157 below demonstrates an alternation between monovalent *asrh* 'to pour out by self' and the derived causative stem *w-asrh* 'for someone to pour out something'. Example 157a is monovalent with the u-prefix in the person/number invariant middle form. Example 157b on the other hand shows the causer of the action identified by ergative case marking, and the morphology indexing the causer on the a-suffix. The undergoer of the pouring action (the water) is indexed on the u-prefix.

(157) a. nu n\asr/ø
water.ABS 3sg:pfv.pst/pour.out
The water poured out.

Nmbo Dictionary entry *asrh*

b. nu y\asr/n, yndo.
water.ABS 1sg.A>3sg.U:pfv.pst/pour.out 1sg.ERG
I poured the water out.

Nmbo Dictionary entry *wasrh*

4.8.4.2 Benefactive/Applicative Constructions

Benefactive/Applicative constructions introduce a beneficiary argument. The diathetic typically takes the form *wa-*, and the additional argument is indexed on the u-prefix. In the example below 158a shows the benefactive/applicative *(w)a-* on the verb root, with the beneficiary cross-referenced on the prefix and expressed by the free pronoun with the benefactive case marker *=wt*. In example 158b there is no benefactive/applicative prefix on the verb, and the free pronoun of the recipient is case marked as a human goal.

(158) a. ymo tärvär svälnḡ w**a-**ramo/i
3sg.ERG plenty basket.ABS 3sg.A>1sg.U:pfv.pret/**BEN.APP**-give
ta=wt
1sg.DAT=BEN
She gave me lots of baskets.
Fieldnotes 2017 Book 2:9

b. ymo tärvär svälnḡ e-n\ramo/i ta=vav
3sg.ERG plenty basket.ABS 3sg.A>1sg.U:pfv.pret/give 1sg.DAT=GOAL
She gave plenty of baskets to me.
Fieldnotes 2017 Book 2:9

4.8.5 Valency Decreasing Constructions

4.8.5.1 Detransitivising Constructions

Detransitivising constructions in Nmbo has the diathetic prefix on the verb taking a vowel form, usually *a-* or *ä-*. The A argument of a transitive verb becomes the sole argument S of an intransitive construction. The ergative-marked A nominal takes the absolutive form. A similar process has been described for Nama as the *antipassive* (Siegel 2017:130-132).

(159) a. nu**-ama** ynd w\vam/**nd**.

rain**-ERG** **3sg.A**>1sg.U:pfv.pst/rain

The rain rained on me.

Nmbo Dictionary 2017 version entry *famh*

b. nu n**a**-vam/**nd**

rain.ABS **3sg.A**:pfv.pst/**DTRN**-rain

It rained. [lit. Rain rained.]

Nmbo Dictionary 2017 version entry *avamh*

(160) a. **yndo** qév y\tan/**an**.

1sg.ERG hole.ABS **1sg.A**>3sg.U:ipfv.nphd/dig

I am digging a hole.

Fieldnotes 2017 Book 1:22

b. ynd n**a**-tan/**an**.

1ABS **1sg.A**:ipfv.nphd/**DTRN**-dig

I am digging.

Fieldnotes 2017 Book 1:22

4.8.5.2 Reflexive/Reciprocal Constructions

Reflexive/reciprocal constructions have the A and U arguments that are the same, i.e. the A argument is simultaneously the U argument, and vice-versa. Morphologically this is manifested by the dual marker and a middle marker prefix, and the A argument is cross-referenced on the a-suffix. The free pronouns are, however, placed in the absolutive. The diathetic is either *a-* or *ä-* as a reflexive.

(161) a. **ynd** bmova n**ä**-gnu/øm.

1ABS 2sg.COM **1du.A**:ipfv.nphd/**RR**-hug

We are hugging each other.

Nmbo Dictionary 2018, entry *ägnuh*

b. kai bä n**ä**-gm/**em**.

± 1day FUT **1du.A**:ipfv.nphd/**RR**-fight

Tomorrow we will fight each other.

Nmbo Dictionary 2018, entry *ägmh*

4.9 Complex Clauses

The description of Nmbo complex clauses is still in the early days. Here I provide brief notes on what I know concerning some complex clause types. These are the phasal constructions, a few complement clause types (subsection 4.9.6), a few relative clause types, and some adverbial clause types.

Nmbo subordinate clauses can be finite or non-finite. The subordinate clause often contains a word that functions as a subordinator, such as the temporal particle *gea* 'if; when' for conditional clauses, and *ka* 'where' for relative clauses.

4.9.1 Phasal Constructions

Phasal constructions (Evans 2015a:544) are clauses made up by two verbs; a nonfinite verb and a finite verb. The finite verbs used in this construction are a limited set, including *wavroh* 'to do' *wivoh* 'to finish', *ramh* 'to make; do', *rñeh* and *wavngoh* 'to begin something'. Impressionistically, it seems that *wavngoh* is the most commonly used phasal verb in the Nmbo corpus. Phasal clauses emphasise the inception of an action (examples 162), and are sometimes translated by Nmbo speakers as "trying to doing X" (example 163). The lexical meaning of X is provided by the nonfinite verb.

(162) a. Szräzr **wavro-h** t-ng**avngo**/n,

house.post.ABS **do-INF** 1sg.A>3nsg.U:pfv.proxin:and/**begin**

.

I started making the house posts.

WSEK1-B20150817-02DimbanKSae, 00:03:36.780 - 00:03:38.390

b. **ram-h** t**avngo**/ngaym.

make-INF 1pl.A>3sg.U:pfv.phab/**begin**

We would all start making [the basket].

WSEK1-G20151015-15BasketMW, 00:03:12.338 - 00:03:16.365

(163) ["Trying to X" phasal constructions.]

a. tanzo zi **tambar-h** e-ng**wavngo**/yn

1sg.POSSC language.ABS **recognise-INF** 1sgA>3nsgU:pfv.pret/**begin**

e~ kor mna mananda fivi n\äyätu/yn

DISC again DEM.fw familiar INTS 1sg.A:pfv.pret/become

I was trying to recognise my language until~I became familiar with it again.

WSEK1-B20170627-03SmakoSL03MQ, 00:04:21.901 - 00:04:24.313

The finite and nonfinite verbs of a phasal construction will typically be placed contiguously. There are, however, examples where another element such as a demonstrative can occur between them.

(164) a. Ge t\rnu/tawm, **ram-h** däv t**avngo**/ngaym,
DEM 1nsg.A>3sg.U:ipfv.rmpst/fill **make-INF** then 1nsg.A>3sg.U:pfv.phab/**begin**
svälnḡ.
yam.basket.ABS
Then we would start making the yam basket.
WSEK1-G20151015-15BasketMW, 00:03:16.365 - 00:03:19.413

b. Tande hkv mñutyu **rar-h** ynao w-n**avngo**/.
1sg.POSS eye.ABS all rar-h DEM.LOC 3sg.A>1sg.U:pfv.pst:ven/**begin**
My eyes were beginning to hurt [lit. My eyes were beginning to be bitten (by the salt)]
WSEK1-A20150917-1SinkingBA, 00:12:20.770 - 00:12:23.917

4.9.2 Conditional Clauses

The basic conditional is formed by the preverbal particle *gea* 'if'; when' in the if-clause. The semantics of *gea* can be either a conditional 'if' or a temporal frame 'when', as is common for many languages (including Papuan; Thompson, Longacre, and Hwang 2007:257). A conditional interpretation is put forward when the predicate of the apodosis is indicating a future time frame (examples 166a and b with a future perfective TAM).

(165) a. [Yayam is talking about how and when they feed a captured pig of theirs.]
Gea n-ng\arend/am, yao ä t\aramo/ngm.
COND 1pl.A:ipfv.nphd:and/wander NEG 3ABS 1nsg.A>3sg.U:**pfv.fut**/give
If we are out and about, we will not give [food] to it.
WSEK1-B20141007-PigYS, 00:15:35.186 - 00:15:36.806

b. [Yavs is talking about the coconut tree he planted.]

Ynd	kr	**gea**	w/m,	tande	tog-am
1ABS	death.ABS	**COND**	1sg.U:nphd/be	1sg.POSS	child-ERG

yna	ag	t\aware/nge	däv.
DEM.prox	coconut.ABS	3du.A>3sg.U:**pfv.fut**/own	when

If I die, my two children will own this coconut.

WSEK1-B20170726-01Yavs00Coconuthm, 00:05:18.470 - 00:05:20.860

4.9.3 Reported Speech

Reported speech in Nmbo is always direct speech. There are two strategies to introduce reported speech. One is to have an inflected verb of speech preceding the quoted speech, such as *mndh* 'to tell' (example 166a), and *owavh* 'to speak'. The other strategy is to use the manner demonstrative *dena* or *dena de* (b), which is translated by Nmbo speakers as "like this". It is common to see both strategies used together, with *dena* following the predicate of speaking (c). The reported speech is often in a separate intonation unit with a pitch reset at the beginning.

(166) a.

t\mndo/n		Namai,	"Yna
1sg.A>3sg.U:pfv.proxin/tell		Namai	DEM.prox2sg.A>1sg.U:leave.alone

q\vra/ngo.	Bm wa	no/tan."
2ABS	must	2sg.U:nphd/walk

I told Namai, "You leave me here. You must keep walking."

WSEK1-B20150804-01SnakeBiteTY, 00:09:05.800 - 00:09:09.800

b. Beverli-m dena **de**, “Yao. Tande toge ämb
Beverly-ERG DEM.mn **ALR** NEG 1sg.POSS son QNT
nmbovi yao ym.”
something NEG 3sg:nphd/be

Beverly was like, “No. I don’t have any sons.”

WSEK1-B20150909-MissingMoneyMZ, 00:05:43.047 - 00:05:45.881

c. Qnze nambt n\owav/t **dena**, “mé tm,
Qnze ±2days 3sg.A:ipfv.nphd/speak **DEM.mn** CONT 3sg:ypst:be
yndo ä y\vene/tan.”
1sg.ERG FUT 1sg.A>3sg.U:ipfv.nphd/feed

A couple of days later Qnze said like this, “Leave it, I will feed it.”

WSEK1-B20141007-PigYS, 00:05:52.948 - 00:05:57.165

4.9.4 Contrastive Constructions

Contrastive constructions are marked by the words *wndede*, *wnde*, or *ede*. These words can occur between two clauses and function as a disjunctive coordinate clause (example 167a), but can also occur elsewhere. In example 167b we see *wnde* occurring left of the contrastive clause predicate.

(167) a. [Nne q\aramo/nge...] **ede**, [tavav ge=tmaron,
food 3nsg.A>1sg.U:pfv.fut/give **but** 1sg.GOAL TOPIC
mnde-vna tmaro yna nne.]
tasty-PRIV 3sg:rmpst/be food

They gave me food... but, to me this food didn’t taste good.

WSEK1-G20150902-03PigGS, 00:13:43.920 - 00:13:49.430

b. [Hom speaking about how Namna is not being spoken by young people any more, while their use of Nmbo is increasing.]

Namna	zi	y\ehon/at.	Nmbo	zi
Namna	language	3sg.A:ipfv.nphd/descend	Nmbo	language

wnde n\armbo/t.

but 3sg.A:ipfv.nphd/ascend

The Namna language is going down. The Nmbo language, however, it is going up.

WSEK1-B20150813-03PerceptualHZ, 00:05:03.707 - 00:05:08.174

4.9.5 Adverbial Clauses

Nmbo adverbial clauses semantically modify the predicate of the matrix clause, by providing additional information on time (section 4.9.5.1), and purpose (4.9.5.2).

4.9.5.1 Temporal

Temporal adverbial clauses frame the time at which some event occurred, or will occur. Subordinators include *sam* 'when', *ge* (adverbial demonstrative), *däv* 'then'. *Sam* and *ge* occur left of the verb in the adverbial clause, but *däv* seems to prefer clause final position (example 168c). In addition, adverbial clauses with the subordinator *däv* seem to occur after the main clause.

(168) a. [**sam** n\unge/aend, ärgv tndo=n...]$_{ADV}$

when 3du.A:pfv.pret/stand.up salt.water.creek bank=LOC

[okei, säläme k-ng\uvträr/t.]$_{MAT}$

DISC shirt 3du.A:ipfv.ypst/take.off

When they stood on the bank of the river... they took off their feathers.

WSEK1-G20170723-01GwN10KK, 00:01:05.029 -00:01:09.109

b. [**ge** n\ovar/ym Govav=et...]$_{ADV}$ [mngo kitong
DEM 1nsg.A:pfv.pref/arrive Govav=ALL house large
yände dena t/maro.]$_{MAT}$
3sg.POSS DEM.mn 3sg.rmpst/be

When we arrived in Govav, there was somebody's big house.

WSEK1-G20140922-MarriageMGVid , 00:02:49.962 - 00:02:53.830

c. [evh yägu t-n\vaun/awng]$_{MAT}$ [ä w\ingo/ynd
sun/day morning 3sg.A>3sg.U:ipgv.rmpst/sun.rise FUT 3nsg.A>1sg.A:pfv.pst/spot
däv kt äi]$_{ADV}$.
when DEM.dist DISC

The day was dawning and that's when they saw me there.

WSEK1-B20150909-MissingMoneyMZ , 00:25:23.654 - 00:25:29.089

4.9.5.2 Purpose

Adverbial clauses of purpose and intention can employ both non-finite and finite predicates. In example 169 the non-finite verb is marked with an allative *=t*, but it is likely capable of taking other locational case markers. The subordinator of purpose is typically the manner demonstrative *mna*, which is additionally marked with locational case markers such as the allative *=t* (examples 169, 170).

(169) [Non-finite adverbial clauses of purpose.]

a. [är t-ng\auwano/nga ta]$_{MAT}$ [**mna=t**,
people 3pl+.A>3sg.U:pfv.fut:and/gather.selves 1sg.DAT **DEM.fw=ALL**
rsa-h=t...]$_{ADV}$
carry-INF=ALL

People gathered themselves to carry me.

WSEK1-G20150902-03PigGS , 00:10:54.190 - 00:10:56.210

(170) [Finite adverbial clauses of purpose.]

a. [t-n\aito/tao yna mrz]$_{MAT}$ [**mna=t**
3sg.A>3sg.U:ipfv.rmpst:ven/follow DEM.prox girl **DEM.fw=ALL**
ge t-ng\ane/ngai]$_{ADV}$
DEM 3sg.A>3sg.U:pfv.phab:and/take

She followed after that girl in order to kidnap her.

WSEK1-G20140922-MarriageMGVid , 00:02:49.962 - 00:02:53.830

4.9.6 Complement Clauses

Nmbo has both finite and nonfinite complement clauses. The matrix clause predicates are single argument predicates. In Nmbo, clauses can form the complements of predicates of wanting/desiring, learning, telling, and possibly seeing.

4.9.6.1 Complements of wanting/desiring

The main verb is a copula, and the complement clause can be either intransitive or transitive. The entire clause right of the copula is the complement, with an obligatory modal particle *gea* 'if; when' on the left of the finite complement predicate. For intransitive complements, the sole argument agrees in person and number with the sole argument of the main clause. For transitive complements, there are examples where the sole argument of the main clause agrees with the person/number of the a-suffix of the finite complement predicate (example 171b).

(171) a. Ynd mende w\m, Govav=t gea w-ng\m.
1ABS desire 1sg.U:nphd/be Govav=ALL SUP 1sg.U:nphd/go

I want to go to Govav. [lit. I desire to go to Govav]..

Fieldnotes 2017 Book 1, p.10

b. Yna **är sombwi** mende e/ren, kkv

DEM.prox **man.ABS two** desire 3du.U:nphd/be garden.ABS

gea y\aro/**wt**.

SUP 3du.A>3sg.U:ipfv.nphd/clear.out

These two men want to clear out the garden. [lit. These two men desire that they clear the garden.]

Fieldnotes 2017 Book 1: p.10

4.9.6.2 Complements of knowledge

Knowledge predicates combine a nominal such as *mwyat* 'know' or *mavña* 'ignorant' followed by a copula. In Nen it is attested that the complement can be introduced by an interrogative pronoun (Evans pers. comm.), but the most common examples in the Nmbo corpus seem to be the use of the completive particle *de* (example 172a) or the manner demonstrative *dena*, or the common collocation of *dena de* (b). Recall that *dena de* often functions as a quotative. There are also examples of *drandmae* 'how' occupying the position of complementiser.

(172) a. ynd mwyat q/maro [de... mato rokar yna-ma

1ABS know 1sg.U:rmpst/be ALR perhaps thing DEM.prox-ABL

bä ko\gamne/nga...]$_{COMP}$

?3ABS 3sg.A:pfv.fut/come.through

I knew that some thing might come through here...

WSEK1-G20150902-03PigGS, 00:02:24.720 - 00:02:30.710

b. ynd mavña q/maro [dena de, dena yam

1ABS ignorant 1sg.U:rmpst/be QUOT DEM.mn event

bä n-ovro/t kt.]$_{COMP}$

FUT 3sg.A:ipfv.nphd/occur DEM.dist

I didn't know that such an event would happen there.

(WSEK-B20150721$_0$1marriageAT , 00:10:10.833 - 00:10:12.917)

4.9.6.3 Complements of seeing

The predicate of the main clause is a divalent verb *wingoh* 'to spot/catch sight of', or *wakeh* 'to watch'. The complement can be a nominalised verb with a locative case marking on it (example 174a), although there are also examples of verbless complements such as in 174b.

(173) a. wén snta y\ingo/yn, naifa-e
tree.branch sharp 1sg.A>3sg.U:pfv.pret/catch.sight knife-DAT
nv-h=mn.
cut-INF=ORI
I found a sharp stick cut with the knife.
(WSEK1-A20150917-1SinkingBA, 00:14:46.567 - 00:14:48.300)

b. yndo y\ingo/sn yände dmave
1sg.ERG 1sg.A>3sg.U:pfv.pst/catch.sight 3sg.POSS wife
biskar bés=an.
cassava fire=LOC
I saw your wife cooking cassava on the fire.
Fieldnotes 2017 Book 1: 9

4.9.6.4 Complements of learning

Complements of *awatambneh* 'to learn' and *waovaroh* 'to teach' are non-finite, with a nominalised verb providing the lexical semantics of the complement.

(174) a. Ynd bva nawatambnetan [ingin freng-h=t]
1ABS just 1sg.U:nphd/learn engine fix-INF=ALL
I learned to fix the engine.
Fieldnotes 2017: 9

b. tanzo-ye mtar-e ä nawatambnetan [ag
1sg.POSSC-DAT slow-DAT FUT 1sg.U:nphd/learn coconut
armbo-h]
ascend-INF

I will teach myself slowly to climb a coconut tree.

(WSEK1-G20150903-03ClimbingGZ, 00:01:57.480 - 00:02:00.428)

4.9.7 Relative Clauses

Nmbo relatives clauses can be described as *adjoined*. Adjoined relative clauses are "typically marked as subordinate in some way" (Hale (1976):78) to the matrix clause, but not by embedding. The relative clause occurs outside of the matrix clause NP (Andrews 2007:214), and in Nmbo the relative clause can occur both to the left or right of the matrix clause. Relative clauses are marked with a variety of relativisers, including interrogative pronouns. As Siegel (accepted) has noted, this strategy of using interrogatives as relativisers is essentially restricted to European languages (Comrie 2006: 136) and languages that have been in contact with them, and now seemingly, the Nambu branch languages.

Relative pronouns are, however, optional. The relative pronouns used include *ge* (adverbial demonstrative section 4.3.3.2, example 175a), *ka* 'where' (example 175b), *emo* who.ERG and *eve* who.ABS (example 176a, 177a).

Nmbo relative clauses schematic:

$$[\text{NP (VP)}]_{MAT}\ [\text{(NP) (REL.PRO) V}]_{REL}$$

(175) a. [yna site är markai **ge** tnmaro]$_{REL}$
DEM.PROX white man.ABS white.fella **DEM** 3sg:prim/come
n\novar/yng Sigabadru hamba=t.
3sg.A:pfv.pret/arrive Sigabadru village=ALL

This white man that came arrived at Sigabardu.

WSEK1-G20150826-03AusWorkZG, 00:04:06.430- 00:04:15.545

b. novaryng kt=ano [niñi **ka** tmaro...]$_{REL}$

3sg.A:pfv.pret/arrive DEM.DIST=LOC demon where 3sg:rmpst/be

She arrived to where the demon I was.

Fieldnotes 2017 Book 1: 75

Nmbo has both agent and object relativisation. In cases of agent relativisation the actor of the matrix clause is also cross-referenced as the agent of the relative clause. In example 176a the pronoun *b¨a* is the third person absolutive which does not provide number information, but the copula *ym* shows agreement with a 3sg argument; the interlinear gloss for the pronoun is therefore glossed '3sg.ABS' to show the cross-referencing of the matrix clause argument on the a-suffix of the relative clause predicate. Note in 176b the matrix clause is verbless, and the core argument is not explicitly expressed as plural[12].

(176) a. **bä** bä ym [bende mngo emo

3sgABS FUT 3sg.U:nphd/be 2nsg.POSS house who.ERG

t\räm/**et**.]$_{REL}$

3sg.A>3sg.U:ipfv.ypst/build

He is the one who built your house.

Fieldnote 2017 Book 1: 9

b. yna är-m [emo-vem mamwi t\gme/tat.]$_{REL}$

DEM.prox man-ERG who-ERG.nsg pig.ABS 3pl.A>3sg.U:ipfv.ypst/kill

These are the men who killed the pig.

Fieldnotes 2017 Book 1: 75

[12]This elicited example did not produce the plural ergative suffix *-vem* as one may expect.

c. tande toge-toge-am [ge t-ng\ne/**tawt**

1sg.POSS child-RDP-ERG DEM **3pl.A**>3pl+:ipfv.rmpst/eat

yna nne...]$_{REL}$ mnd k-ng\aweto/**tawt**

DEM.prox food vomit **3pl.A**:ipfv.rmpst/masticate

My children that ate this food, (they) vomited.

WSEK1-A20160720-07DroughtWZ 00:01:17.183 - 00:01:21.085

In **object relativisation**, the undergoer of the matrix clause is cross-referenced in the relative clause.

(177) [Example (a) with a relative pronoun *eve* (who.ABS), (b) without relativiser.]

a. yndvem **ew**\ake/tam toge-toge [otambe-h=an

1nsg.ERG 1nsg.A>**3pl.U**:ipfv.nphd/watch child-RDP race-INF=LOC

eve n\rs/**at**.]$_{REL}$

who.ABS **3pl.A**:ipfv.nphd/run

We are watching the children who are running a race.

Fieldnotes 2014: 30

b. yndvem **ew**\ake/tam [toge-toge n\akmw/**et**.]$_{REL}$

1nsg.ERG 1pl.A>**3pl.U**:ipfv.nphd/watch child-child **3pl.A**:ipfv.nphd/swim

"We are watching the children who are swimming."

Fieldnotes 2014: 30

Part III

Quantitative Descriptions

Chapter 5

Methodology

> "...the reader can judge to what extent [the data] gives good representation of the various sections of the population..." (Labov 1966:155)

The Kerake lands lie some 580 kilometres as the crow flies from Port Moresby. The physical distance between the two locations reflects the lifestyle distances between the capital of Papua New Guinea and the remote Nmbo speaking community. As the previous ethnographic and qualitative chapters have shown, life in the Morehead district retains many characteristics of pre-colonial and pre-national society. Working in these communities, while rewarding, brings certain methodological challenges. Matching the quality of quantitative data to the standards set by contemporary qualitative LVC studies makes the entire process additionally challenging (cf. Chirkova, Stanford, and Wang 2018, Meyerhoff and Stanford 2015). Mansfield and Stanford (2017) have outlined six challenges of doing LVC studies in lesser-studied indigenous communities, all of which apply to documenting sociolinguistic variation in the Morehead Area. These are the shallowness of prior linguistic documentation, linguistic structural divergence from world languages, under described communities from a broader LVC perspective, cultural distinctness, and difficulty of entry into the community. These have all played a role in this project, from data collection through to analyses.

This chapter exists to lay out the richness and limitations of the data in light

of the challenges of documenting variation in remote Papua New Guinea. I will provide details on the creation and composition of the two corpora used in the quantitative studies of this thesis within this context of doing socio-cultural and linguistic documentation concurrently. I discuss the data collection processes, and the parameters of the data itself. The Nmbo Wordlist Corpus (NWC) was used for the vowel sociophonetics study (chapter 6). The Nmbo Sociolinguistic Corpus (NSC) was complied to investigate multiple linguistic variables, of which two are presented in this thesis: word initial [h]-drop (chapter 7), and the emergent topic marker (chapter 8). Content of the NSC, such as speaker language biographies and attitudinal data, contributed concrete quotations to the qualitative descriptions (chapters 2, 3). Towards the end of this chapter I will also present a qualitative description of the statistical methods used to investigate the data (section 5.4).

As mentioned in the introduction of the thesis, the choice of linguistic variables was broadly based on opportunity and feasibility. The intention was to study at least one phonetic-phonological variable, and one morpho-syntactic variable. The final set of variables studied are vowel sociophonetics (chapter 6), word initial [h]-drop (chapter 7), and a grammaticalisation of a topic marker (chapter 8). More precise rationales for each variable is provided in each chapter.

Both corpora were composed in conjunction with ethnographic and linguistic documentation. Fieldwork was conducted across the years 2014 - 2017, for a total of eight months. The trips ranged from two weeks to three months in duration.[1] The data collection was done together with a locally created language committee, whose members are listed in the acknowledgements of this thesis.

The design principle behind both corpora was to sample licensed speakers of Nmbo. The intention was to include speakers with the following characteristics: 1) the speaker was primarily raised in a community of other licensed Nmbo speakers, and 2) the father of the speaker is also a licensed Nmbo speaker. The speakers, in both corpora, were instructed by another licensed Nmbo speaker to respond in Nmbo. More detail about the methods of instruction and elicitation are provided in the respective sections of the two corpora in this chapter (sections 5.2 and 5.3.1).

Table 5.1 summarises the data source, token counts, and quantitative methods

[1] More precisely: six weeks in Sept.-Oct. (2014), three months and two weeks in July - Oct. (2015), two weeks in July (2016), and two months and two weeks in June - Aug. (2017).

employed for the three studies.

	Vowel sociophonetics	Word initial [h]-drop	Topic marker grammaticalisation
Data Source	Nmbo Wordlist Corpus	Nmbo Wordlist Corpus + Nmbo Sociolinguistic Corpus	Nmbo Sociolinguistic Corpus
Speakers (n)	63	50 (NWC) 38 (NSC)	38
Tokens (n)	Various	15 per speaker (NWC) 1649 (NSC)	513
Quantitative Methods	Generalised linear mixed models	Generalised linear mixed models	Frequency counts

Table 5.1: Summary table of data source and methods by quantitative study.

5.1 The Role of the Speech Community in Data Collection

A few notes should be made about the role of speech community members, and more specifically, the role of the Kerake Language Committee (KLC), in the process of data collection. I do this to be transparent about the somewhat skewed composition of the speakers represented in the corpora, and will leave the reader to decide what this means in terms of the nature of the data being used. This section is also an acknowledgement of community support and involvement in the creation of the corpora.

The KLC was a locally created entity who greeted me in my first year of fieldwork, which suggests the group that was formed locally with some level of wider community endorsement. I say "some level", since during the course of my fieldwork I would hear the rare voices of dissatisfaction that the KLC was formed without broader consultation across all three villages of Govav, Bevdvn, and Arovwe. Indeed, it turns out that the initial members of the KLC were men who were all tied by kin affiliation at their grandparents' generation. From the outset the KLC appeared to be a balanced representation of social groups: there were four men each from Arovwe and Bevdvn, and two men from Govav. The two men from Govav represented the Mayawa section, while Arovwe was ostensibly represented by Bangu section men, and Bevdvn by Sangara men. But in an area where immediate kin ties are one of the most powerful relationships between individuals, the exclusion of non-kin from the original KLC suggests that some local politics were at play, to which I was not and am still not privy. In other words, my reliance on the KLC to reach out to speakers means the speakers represented in the corpus are highly likely to be members of a particular social network.

The composition of the KLC changed throughout the course of my fieldwork. In my first year we negotiated the involvement of women in the KLC, and gender parity at least in terms of number of female committee members was achieved. In the second year, Govav men with looser kin affiliations to the initial KLC members chose to join the committee. Some members rotated over the course of my fieldwork, but others remained constant throughout the total four years.

Given all this, it should be borne in mind that the interviewers of the two Nmbo corpora also bring unknown effects to the elicitation of speech. In sociolinguis-

tics more broadly, a great deal of attention has been paid to the effects of the observer's paradox and how it can affect the production of a person's speech away from the vernacular. Attempts were made to minimise the effects of the observer's paradox by choosing appropriate community insiders as interviewers (e.g. as done by Cukor-Avila and Bailey 2001, c.f. Wilson 1994), but it is worth noting that the personal histories and relations between KLC interviewers and their interviewees can also potentially affect the interview. An assumption was made that matching age, gender, and village of residence would put the interviewee at ease, ergo most likely to speak freely without monitoring. This is, however, an untested assumption. In a small community where most people know one another, it is possible that some effects unknown to the outside linguist were at play (Baclawski 2018:79 for role of community interviewer in Cham, Vietnam, Abtahian 2018 for Garifuna, Belize). For example there may be past grievances, or a clash of personalities that were unknown to me at the time.[2]

Finally, the vast majority of recordings were done in open spaces in the presence of other community members. The KLC and other community members deemed it inappropriate to sequester speakers away from the eyes of community members, and better for the outsider to be open and transparent in the work they were conducting in the community. Thus most recordings were done where the target speaker was speaking to the interviewer, with curious community members watching on and occasionally prompting the target speaker to give certain answers. Whether this public form of speaking approximates the vernacular, or whether it is a genre unto its self is up for further investigation. Public speaking and narration in the presence of community members is a common speech act, which provided speech that is consistent with the goals of sociolinguistic interviews that prioritise naturalistic and free-speech over elicited and highly-structured speech.

A handful of interested and linguistically astute community members, from within and outside the KLC, were involved in the transcription and translation of recorded speech. Transcribers were reimbursed financially for their time and knowledge, and received basic training in the use of laptops as part of the work. Some community members were cultural consultants who worked along

[2]For example one of the interviewers from Govav was, I found out later, quite disliked by certain Bevdvn and Arovwe KLC members. No specific reasons were given. I have found in my experience that the Kerake people are quite private, and rarely complain about others in a group context. The interviewer from Govav only ever interviewed fellow villagers from Govav, but I do not know whether he was disliked as much by other Govav villagers.

side the community members who spoke English to translate the contents of recordings, and many transcribers would directly transcribe the Nmbo orthography into ELAN (MPI, 2018), a specialist software for linguistic transcription and annotation. All but one transcriber were first-time computer users, and by the fourth field season there were two teams of two who could transcribe in ELAN independently. A full list of community members who were involved are listed in the acknowledgements.

5.2 Nmbo Wordlist Corpus (NWC)

The NWC is a corpus consisting of three different word lists collected across the four years of fieldwork. It is a corpus that evolved along with the documentary project of Nmbo, and so has a somewhat eclectic collection of words and speakers. The words in the list vary across the four years, but the 2016 and 2017 word lists are identical. The full word list is presented in table 5.2, and the break down of words sampled by year is presented in the appendix (A.16).

Speakers were not entirely chosen at random, since this is challenging when the total population of Govav and Bevdvn is roughly 400 people. Individual speakers were approached based on opportunistic availability, but attention was paid to sample across ages, genders, and villages of origin. An attempt was made to sample the same speakers across the four years with mixed success. The single consistency concerning speakers across the four years is that a minimum of 17 speakers have been sampled from Bevdvn, with at least seven of those being women.

Speakers were asked to repeat the target word three times in isolation without a carrier phrase. Since most members of the speech community are non-literate, the word list was administered verbally. The verbal administration of the word list is treated as analogous to reading an orthographic representation of target words. Almost all the target words are lexical nouns, with the exception of *bm knm* (imperative 'you come') and the non-finite verbs *wuvuh* ('to check/scout') and *itroh* ('to move'). I would say the word once in English then in Nmbo, and the speaker would repeat. For [h]-words which were known to vary in their realisations (i.e. *hamba* vs *amba*), the linguist used the more conservative [h]-ful pronunciation. All speakers, across all four years, were recorded with a Samson SE10 head-mounted microphone at a sample rate of 44.1kHz on a Zoom H2n recorder. The recordings were usually done outdoors in the presence of other

community members.

The opportunistic way in which this data were collected has meant that unfortunately there are major gaps in the representation of certain words and speakers across the four years (A.10). This has meant that the statistical analyses conducted in the vowel sociophonetics chapter deals with aggregates for individual speakers and phonemes across various years, and finer grained effects of word are obscured. It would have been preferable to go into the field equipped with a well-controlled word list, and an ability to approach specific individuals to meet the desired number of speakers per social categories investigated. Nonetheless, the data reveal some interesting patterns, as will be presented in chapters 6 and 7.

For the acoustic analyses, all vowel tokens were located in the digitised waveform by the Munich Automatic Segmentation System (MAUS) (Kisler, Schiel, & Sloetjes, 2012). Beginning and end-points were all checked manually and adjusted where necessary. 30 equally spaced values of the first two formants (F1, F2) in the central 60% of each token were obtained in Praat (Boersma & Weenink, 2016). The Burg algorithm was used with formant ceilings of 5500Hz and 5000Hz for female and male tokens, respectively. Each set of 30 values was then fitted with second-order discrete cosine transform curves, which produces smoother formant trajectories (Williams & Escudero, 2014). Median F1 and F2 values from the smoothed trajectories were used to represent phonetic vowel quality, which has been shown as efficient for monophthongs vowels (Moore & Carter, 2015).

5.3 Nmbo Sociolinguistic Corpus (NSC)

The NSC was designed to sample speakers of two speech communities: Bevdvn and Govav. The veracity of village as a way of defining a speech community is discussed in section 3, but the key points are as follows. Firstly, Kerake commonly refer to themselves on a village basis, such as *Bevdvnmn* 'originating from Bevdvn', *Govavmn är* 'a person from Govav', perhaps akin to saying 'Sydney-sider' or 'Bostonite'. Secondly, social activities such as church gatherings are often conducted and organised on a village-by-village basis. Thirdly, areas of residence are clustered in villages. While not all Kerake reside in the village at all times throughout the year, the relative permanence of housing indicates that daily life and socialisation is centred around the physical location of the

Vowel	Words
i	bi.træ.r**i**, ka.k**i**, t**i**.kv, w**i**n.goh, f**i**.v**i**, **i**.troh, z**i**.wæl næn.z**i**
e	**e**.vh, m**e**r, son.zær.bw**e**, d**e**.v**e**, d.ma.v**e**, **e**nd, h.ra.r**e**, hu.r**e**, m**e**n, n**e**n, tan.d**e**
æ	**æ**. gæ .vu, bæ, bæ .rm.bæ r, bi.træ .ri, bu.dæ r, dæ v, fæ tr.fæ tr, s.gæ r, son.zæ r.bwe, **æ** mb, svæ .lnḡwæ z, zi.wæ l, n**æ**, n**æ** n.zi
u	æ.gæ.v**u**, b**u**.dær, b**u**.tém, h**u**.re, h.v**u**, w**u**.v**u**h
o	a.d**o**.d**o**, g**o**t, mn.g**o**, s**o**n.zær.bwe, t**o**k, t**o**t, wei.k**o**r, win.g**o**h, b**o**m.b**o**, **o**.rng
ɑ	**a**.do.do, h**a**m.b**a**, h.h.n**a**r, k**a**.k**a**, k**a**.ki, k**a**.n**a**m, s**a**.w**a**, d.m**a**.ve, sr.v**a**v, z**a**v, **a**.nu, f**a**.der
ə	bu.t**ə**m, ḡ**ə**.ḡ**ə**n, s**ə**.gær, b**ə**r.b**ə**r, b**ə**.rm.b**ə**r
ɐ	b**ɐ**m.k**ɐ**.n**ɐ**m, e.v**ɐ**h, h.k**ɐ**v, h**ɐ**.ra.re, h**ɐ** .vu, k**ɐ**.k**ɐ**v, m**ɐ**n.go, q**ɐ**.ki q**ɐ**m.biv, q**ɐ**t, q**ɐ**v, ti.k**ɐ**v, s**ɐ**q , s**ɐ**r.vav, s**ɐ**.væ.l**ɐ**nḡ, t**ɐ**nd, b**ɐ**.rm.b**ɐ**r

Table 5.2: Full set of NWC words.

village.

Since the goal of this thesis is to give an initial characterisation of sociolinguistics of variation in Nmbo, I have opted for what Meyerhoff and Walker call a "long-thin analyses" (Meyerhoff and Walker 2012:217). In contrast to "short-thick" analyses, long-thin analyses work with a smaller number of tokens from a somewhat large number of speakers, with the aim of investigating the social conditioning of variation across the speech community. While the number of speakers represented in the NSC can hardly be characterised as large, the intention behind the design is the same.

5.3.1 Data collection and composition of corpus

This first version of the NSC comprises 15 hours and 56 minutes of continuous speech (i.e. excluding pauses, gaps, and segments with no speech), with 8 hours and 15 minutes worth transcribed. The aim was for each speaker to contribute 30 minutes of continuous speech, but due to the difficulties in getting younger speakers in particular to speak, this was not possible. Unlike the NWC, the NSC was designed deliberately to target a certain number of speakers: a minimum of

nine men and women, with three speakers per age group of young, middle age, and senior. The final corpus comprises 9 women and 11 men from Bevdvn, and 8 women and 10 men from Govav (see tables 5.3 and 5.4). Speakers were recruited through the KLC, and many of the speakers represented in the NSC are members themselves. A speaker database containing additional biographical data can be linked to the NSC, providing details such as clan, mother's emblematic language, language repertoire, additional villages of residence, schooling, and travel.

As mentioned in the introduction, the design principle behind both corpora was to sample licensed speakers of Nmbo that were primarily raised in a community of other licensed Nmbo speakers, have fathers who are also licensed speakers of Nmbo. The profiles of the men in the NSC are relatively straight forward, but there are a few particularities to note about the women. The men chosen for the NSC are all Kerake men who are from, and reside in, a Kerake village. The Bevdvn speakers have a more complicated history due to their migration from Arovwe to Bevdvn in the 1980s, but they self-identify as coming from and living in Bevdvn. Due to the complex nature of the Morehead area and the practice of exogamy, a good portion of the women in the NSC reside in non-Nmbo speaking villages. These women can be characterised as *daily bilinguals* (section 5.5), who report themselves as daily using both their own Nmbo language, as well as the language of the village. Seven out of the 17 women in the NSC have married to non-Nmbo villages, of which six have married to the Nen speaking village of Bimadbn. For the statistical analyses, all these daily bilingual are treated as a single group. Of those, four are from Govav. Many women from Govav married out to neighbouring villages (Rueck, 2006), while this is less the case for Bevdvn. Many Govav women married into the Yarne village of Drdr, but they are not represented in the NSC.

Due to the sample size of the speakers in the corpus, Arovwe residing individuals were collapsed into the category of Bevdvn speakers. As mentioned in the speech community section (3.3.3), the Bevdvn and Arovwe villagers with whom I worked reported they speak the same way, in contrast to the villagers of Govav. The speakers residing in Arovwe are marked in table 5.3 with the notation (A) beside their code.

A number of conversational topics are represented in the NSC. Topics and speech events range between naturalistic and semi-structured formats. By *naturalistic* I mean unprompted monologues without interruptions by questions by the in-

Village of Origin: Bevdvn				
Female	YoB	Residence	Male	YoB
yfbbAY	1990		ymbbME	1997
yfbbAT	1989		ymmmSL	1989
yfabRM	1989		ymbbMZ	1985
yfbbJY	1986		ymbbTT	1982
mfbmYS	1972	Nen village	mmbbKS	1979
mfaaRT (A)	1969		mmbbKS2	1977
sfamRA (A)	1960	Nen village	mmaaBA (A)	1969
sfbaWZ (A)	1957		mmbbLS	1967
sfbfHZ	1949	Yarne village	smbbTS	1953
			smbbYZ	1951
			smbbSZ	1948

Table 5.3: NSC Bevdevn speakers. 9 women, and 11 men. Residence village is Bevdvn unless stated otherwise in column *Residence*. (A) = resident of Arovwe

Village of Origin: Govav				
Female	YoB	Residence	Male	YoB
yfggMW	1999		ymggGW	1999
yfgmBZ	1989	Nen village	ymggBS	1994
mfgbAN	1970		mmggRG	1985
mmggMW	1968		mmggGZ	1984
mfgmGN	1966	Nen village	mmggGS	1976
mfgmGwN	1965	Nen village	mmggMN	1968
sfgmFY	1955	Nen village	smggZG	1959
sfggDS	1947		smggGM	1954
			smggMG	1954
			smggMY	1952

Table 5.4: NSC Govav speakers. 8 women and 10 men. Residence village is Govav unless stated otherwise in column *Residence*.

terviewer. *Semi-structured* speech events involve questions and prompts by the interviewer. All speech events involved a local Nmbo speaker, and the presence of the linguist. In order to reduce the effects of the observer's paradox, Nmbo interviewers were best-matched to the target speaker by gender, age, and village of residence. A break down of tokens per speaker by speech topic is provided in the appendix (A.6)

Naturalistic	Semi-Structured
Personal narratives	Coconut interview
What I did yesterday	Multilingualism questionnaire
Near death stories	Cassowary Picture Task
Ethnographic descriptions	Perceptual dialectology interview
Marriage stories	

Table 5.5: Speech topics covered in the NSC.

The naturalistic topics and speech events are as follows: *What I did yesterday* provides an exposition and summary of what the speaker did the day before. *Personal narratives* and *near death stories* are the same topics outlined in Labov's New York City sociolinguistic questionnaire (Labov 1966:146). Ethnographic speech events are narration of ethnographic knowledge, such as how to make a particular artefact, explanations about flora or fauna, and explanations about social practices. This is akin to localised modules used by variationists adapting Labov's modules to other social contexts (e.g. Meyerhoff and Walker 2012 for the Eastern Caribbean, Rojas-Berscia 2019 for the Shawi of the Amazon). *Marriage stories* were recorded for speakers who had a husband or wife, and had the speaker recount how they came to be married.

A variety of semi-structured interviews were conducted. The types of questions that were asked are listed in the appendix (A.7), but the brief descriptions of these interviews are as follows.

The *Coconut Interview* is a style of data collection created by Evans (2017a), centred around asking speakers about why they planted a particular coconut tree. Coconut tree growth in Southern New Guinea is a result of deliberate planting for a variety of reasons. Not only are the trees useful as a source of nutrition and as a source of material creation, the trees are monuments to autobiographical memory and serve as genealogical investments (Evans 2017a: 29-30). The autobiographical memories that individuals choose to mark are unique, ranging from marking the end of a successful church fellowship, the coming of the new millennium, and even a commemoration of the election of a local member to government. The target speaker explains their rationale for planting a coconut, and the interviewer will proceed to ask about the various uses of the coconut tree and its parts.

The *multilingualism questionnaire* is a combination of biographical interview

with targeted questions about the speaker's multilingual capabilities. The target speakers are asked to report on their language use with their mother, with relatives, and were asked to give a short demonstration of their mother's language if they were willing. Interviewers would often ask specifically whether the target speaker could "speak" or "hear" other languages as a means to interrogate the target the linguistic ability of the target speaker (see section 2.7, table 2.1 for discussion on local expressions of linguistic ability).

The *Cassowary picture task* is a picture-task designed to target h-words, but within a task-based context. The task was inspired by a variety of picture based tasks, such as the Family Problems Picture Task (San Roque et al., 2012), and work by Mansfield (2015) mimicking a rapid anonymous survey technique for Murrinhpatha verbal suffixes. There are two phases in the Cassowary picture task: an activity phase and a retelling phase. Speakers were first asked to describe a series of sequential images, and then asked to retell the story to a group of children using the pictures as a prop. The intention was to have the speaker focus on the image decoding, and retelling performance, that they will monitor their speech less when it came to the h-words. The Family Problems Picture Task identified some minor problems with the interpretation of graphical conventions such as speech bubbles (San Roque et al. 2012:152), but this was overwhelmingly not the case in the Nmbo speech community. The pictures were designed specifically to target Nmbo [h]-words, and illustrated to suit the environment of the Morehead District with appropriate natural features. This makes the interpreting and reporting of statistical significance somewhat problematic, but as we will find in chapter 7, some initial patterns can be revealed which are worth reporting on.

The various speech events and topics were recorded across multiple sessions across four years. Most recordings were both audio and video recorded, using both a head mounted mic recording (Samson Head Mounted at a sample rate of 44.1kHz), and a general recording (Zoom 5, sample rate 44.1kHz).

5.3.2 Corpus Transcription

Unlike the NWC, the NSC has textual mark-ups to indicate discourse events that may affect the realisation of the linguistic variables under investigation (e.g. pauses following word final t/d in New Yorker English triggering deletion in Guy 1980:27, but disfavouring deletion in Philadelphia, Labov 1989:90). A portion of

Figure 5.1: Images of Data collection. Top-left a typical recording set up for narratives. Top-right the cassowary picture task set up. Bottom-left a coconut interview session. Bottom-right transcription work by Mr. Gima Zoga and Mr. Karuwa Wagra.

the NSC recordings were transcribed and translated into English. Coding for the various variables were done in a text-based format. The coding of each variable will be outlines in the relevant chapters, but the general principles of annotation will be outlined here.

The transcription method used is an adapted version of what was developed at the University of California, Santa Barbara (Du Bois, Schuetze-Coburn, Cumming, and Paolino 1993). The transcription method I have used divides streams of speech into intonation units, with attention paid to transitional continuity. The complete Du Bois method involves paying attention to a multitude of factors, but many of these were deemed peripheral for the NSC so were not annotated (e.g. overlaps in speech). Factors such as vocal noises, precise duration measurement, quality of the speakers' voices, non-linguistic sounds, were also not annotated.

Speech is broken down into prosodic units of intonation. An intonation unit (IU) is defined loosely as "a stretch of speech uttered under a single coherent intonation contour" (Du Bois et al. 1993:47). Intonation units in the NSC are marked by three types of transitional continuity: the final stop, two types of continuing markers, appeals, and truncations.

The *final stop* in Nmbo is the same as terminal falling pitch found cross-linguistically. It is marked by a period in the practical orthography. As Du Bois et al. 1993 note, this symbol represents an intonational category rather than a syntactic one (p.54), although in the NSC it appears that these categories overlap very often.

I have marked two types of continuing markers as described in the sketch grammar (section 4.1.4). One is a relatively flat-pitched transition marker, whose transitional continuity is understood as 'continuing'. In Nmbo this *continuing marker* shows a very modest drop in pitch, frequently with little or no pause following it. This boundary marker is indicated in the annotations by the ' , ' symbol. The *rising continuing marker* often shows a rise in pitch which sounds different to the questioning intonation, and is marked in the annotations by ' ... ' . The rising continuing marker is almost always followed by a pause of some duration. Consequently, the ' ... ' symbol is also used indicate that there is a following pause, which is longer than 0.4 seconds.

Below is an example of text marked up with the transitional continuity symbols outlined above. Note that individual IUs do not appear in separate lines in the actual textual corpus, and precise pause durations are also not shown alongside the transitional continuity symbols. The pause duration in parentheses below is presented to indicate the relationship between the boundary marker and approximate pause following it.

yndo ä yaitotan, (0 sec)
qäu gmhmn zi. (2.0 sec)
savat evhon tnmaron, (0 sec)
skul toge-toge nowanoym, Bimadvnat. (1.3 sec)
ynd, (0.2 sec)
Kambora, (0 sec)
yndvem tämséwm... (1.2 sec)

I will tell a story,

about the time of the python killing.
On Sunday at noon, all of us went,
the school children arrived at Bimadbn.
Me,
Kamobra,
the two of us escorted them...

(WSEK1-G20151015-17PythonMW, 00:00:27.685 - 00:00:40.527)

5.4 Statistical Modelling and Interpretation

The goal of using linear mixed models is to quantitatively test the relationship between predictors and interval scale outcomes. In other words, it is a model that tests for correlations between predictor variables and a numerical outcome. Linear mixed models set predictors (or variables) which the analyst believes to have an effect on the patterns of variability, and then runs the model to tests whether, and how much, these fit the data that are provided. The better the fit of the model, the more likely that the model is capturing the distribution of the data.

Recent methodological standards in variationist sociolinguistics the related field of sociophonetics employ linear mixed models as a standard of quantitative investigation (Johnson 2009, Drager and Hay 2012, Baayen, Davidson, and Bates 2008). Linear mixed models include two types of effects: fixed and random. Fixed effects are the variables that are directly being tested, and are (theoretically) replicable in other studies. Common fixed effects in variationist studies include speaker gender, and class. Random effects are the variables that are usually not replicable, such as the effect of individual speakers and their idiosyncrasies. Random effects are included in a model, in a sense to introduce noise to the model so it may better reflect the complexities of variables in real social phenomena.[3] Recently Meyerhoff and Klaere (2017) have raised questions about

[3]Statistical models are necessarily abstractions and simplifications of phenomena, and need to be constructed in a way that approximates the real world in a sound way. A statistical model which includes only a handful of fixed effects will place undue numerical weight on these variable, which is a questionable approximation of real world social phenomena. For example, if a statistical model tests only for gender and age, the numerical input will be spread across these two variables alone. The results will be very strong and might show some numerical/statistical significance, but this result is likely to be unreliable as far as accurately capturing what is responsible for the patterns in the data (called Type I errors in statistical parlance).

the methodological soundness of introducing random effects in small datasets, but for this thesis I have decided to follow current methodological orthodoxy and include speaker and word as random effects.

I model random intercepts, one of the two types of random effects for the statistical tests used in this thesis. Random intercepts and random slopes are two ways in which random effects are accounted for in the data. Random slopes are a numerically more sophisticated way of accounting for random effects in data, but have numerical particularities that need to be met in order to be successfully modelled. The datasets constructed for this thesis were too small to run random slopes models, so random intercepts were used as the next best solution.

Effect size of the variables are reported on using the notation R^2m and R^2c. These are values that explain the proportion of the variance in the outcome variable, which can be accounted for by the predictor variable. R^2m (marginal coefficient of determination) is a measure used for fixed effect models. R^2c (conditional coefficient of determination) is a measure for mixed models which utilise both fixed and random effect (Nakagawa and Schielzeth 2013:136). For example, if a model sets gender and age as two fixed effects with speaker and word as random intercepts, an R2c of 0.15 indicates that the variables/the model explains 15% of the variance in the outcome. There is no clear convention as to what is considered a robust R^2 figure for sociolinguistic studies, but a figure of 0.5 is considered in sociology as a moderate effect (Mooi and Sarstedt 2011:211).

The mixed models were fit using the lme4 package (Bates, Mächler, Bolker, and Walker, 2015) in the R suite (Team, 2017). All test results are deemed significant when the chance probability of the statistic is less than 0.05. R^2 is calculated using the etaSquared function in the lsr package (Navarro, 2015).

5.5 Concluding Remarks

Describing sociolinguistic variation in conjunction with language description is an endeavour which is increasingly viewed as important; both speech communities and the scientific community. This chapter detailed the data, its depth and limitations, and the processes undertaken to obtain these for the creation of the NWC and NSC. The basic ground work has now been laid for the quantitative studies, and in the following three chapters I will present the results of this challenging but highly rewarding process.

Chapter 6

Study 1: Vowel Sociophonetics

Ämb km gs engm fnatae kt ä engm. "Some voices [dialects], ok, they are different." (Kawas Säme, Perceptual Interview 2015)

This study is an acoustic sociophonetic description of the Nmbo vowel space. The goal is two-fold. One is to identify some basic acoustic characteristics of the Nmbo vowel space. The other goal is to identify and analyse social variables of the speakers that explain the patterns of variability in the vowel space. The acoustic measures of vowel realisations investigated are F1 and F2, with the former measuring vowel height and the latter vowel fronting/backing. Vowel duration, and vowel space spread, are also investigated.

Standard variationist methodology would choose to investigate highly variable linguistic phenomena. This study was an experiment in investigating a part of the language that appears relatively invariable, but is known to vary cross-linguistically. An associated aim was to see whether there is any latent or nascent microvariation that are conditioned by sociolinguistic factors.

This study is a concrete case of concurrently doing description and studying variation. This study utilises statistical modelling (section 5.4) in a slightly unconventional way for these ends. Linear mixed modelling is employed less for hypothesis testing, and rather for capturing the conditions that affect the patterns of variability in the vowel space. The statistical models are generally fragile, but there are some promising signs of interesting patterns. Overall we

find a stable vowel space with some micro-variations in vowel realisations, with speaker age and village of origin affecting different vowels and different measures. Speaker and word also play a major role in explaining the patterns of variation, but the variable of gender is not systematically significant.

This chapter begins with a short overview containing the rationale of conducting a sociophonetic study (section 6.1). I then will outline additional methodological notes to supplement the main methodology chapter 5 (section 6.2). Next I will present the results of statistical modelling (analysis of variance models (ANOVA) or linear mixed models) that show systematic social conditioning in some parts of the Nmbo vowel space (section 6.3). I will conclude the chapter by contextualising the findings in terms of the details of the Nmbo speech community (section 6.4). The relevant sections of the sketch grammar for this study are the phonology section (4.1), and in particular the vowel section (4.1.2).

6.1 Overview and Rationale

Phonetic studies of Papuan languages are rare, and sociophonetic studies even more so. Acoustic phonetic studies of Papuan languages have tended to be for basic descriptive purposes, such as focussing on the consonants and vowels of the phonemic system (e.g. Evans and Miller 2016 for consonants and vowels of Nen, Cottet 2015 on Mbahám, Steed and Hardie 2004 for consonants of Kuman (Chimbu, Trans-New Guinean)), or on prosody or tone (e.g Lindström and Remijsen 2005, Odé 2002, Donohue 1997). Exceptions to these foci are rare (an exception being Cottet 2014 on assibiliation processes in Mbahmám). The first sociophonetic vowel acoustic study conducted on a Papuan language was by Kashima, Williams, Ellison, Schokkin, and Escudero (2016), investigating the Nmbo vowel space of nineteen speakers from the village of Bevdvn. The study in this chapter expands the sample to include speakers from the village of Govav, making this study a larger investigation into the wider speech community of Nmbo speakers.

The term *sociophonetics* is a relatively new term, with Foulkes and Docherty (2006) tracing the first use to Deshaies-Lafontaine (1974). The term was used primarily by phoneticians in the late-20th century to mean "descriptive accounts of speech production across different dialects, speaker groups or speech styles" (Foulkes and Docherty 2006:412). Much sociophonetic work in recent years has focussed on the role of social indexical meaning and style shifting, and the con-

struction of identities through phonetic means; what Eckert (2012) famously calls the *third wave of variation*. There are a wide range of sociophonetic studies that look at constructions and negotiations of identities in a variety of social context (e.g. Zhang 2005 yuppies and tonal phonology in Beijing, Mendoza-Denton 2008 Latina gangs and phones in Californian English, Podesva 2011 gay men and declarative contours in North American English, Drager 2015 teenage girls and consonants, vowels in New Zealand English). These studies argue that synchronic variation is a reflection of how speakers utilise phonetic-linguistic features for symbolic purposes. In this Nmbo study, however, I investigate acoustic differences and similarities between speakers in a more "first wave" and "second wave" form. As Eckert herself points out, the concerns of these early variationist studies were to trace the spread of linguistic change through social space (2008: 88). The goal for the Nmbo investigation is to understand which social categories are manifested in the vowel space, and what these social categories may reveal in terms of the social space and configuration of the wider speech community. Nmbo vowels are particularly well suited for this end, since vowel differences are not a salient point of difference when discussing linguistic differences across age, gender, or the Other Village.

One way in which the social space of a community is revealed, is to view the similarity of pronunciations between speakers and groups as an indication of sustained, shared interactions in the past. Accent similarity is understood as shared communication and interaction between geographically congregated people (e.g. Denes and Pinson 1963), and theories such as exemplar theory propose that knowledge of linguistic structures are built up in the memory of an individual over the totality of linguistic experiences they have had (Pierrehumbert 2001, Johnson 2006, Lacerda 1995). Similarities in speech between individuals would therefore be due to sustained input and production within a communication network. Studies focussing on human psychology also suggest that there are social motives for accent similarities. Communication accommodation theory places social motives as the drivers of accent convergence between individuals and among groups (Giles and Ogay 2007:297, Giles and Coupland 1991). Cohen (2012), in a summary paper reviewing the literature on the evolution of human cooperation via accent similarity, finds compelling evidence that accents are salient guides to social preferences in building trust, prolonged interaction, and cultural learning. Evidence from child language acquisition shows that preschool children match and modify their dialect features from primary caregivers to peers as they become

socialised into the wider community (c.f. Locke 1993). Variationist work has also found that people continue to adapt their accents to their social situations during the course of their life time (Sankoff and Blondeau 2007, Rickford and Price 2013). This is all to say there is evidence from a variety of studies which suggest humans make and detect fine-grained similarities in speech acoustics with the the people they interact with commonly. Exposure to interlocutor and community speech acoustics in turn builds up speech community norms. As Pierrehumbert (2006) states, there is a "circular causality of population-level feedback" (p.517) between phonetic variation and accent similarities within members of any given social group.

Sociophonetics of small-scale communities are becoming more common, but many of these are in contexts of contact with a larger regional variety or national language (e.g. Clarke 2009 on Sheshatshiu of Canada, Rau, Chang, and Dong (2009) on the Yami of Taiwan). Vowel acoustics studies of small-scale communities that are relatively free of such contact situations are very rare (for an exception, Stanford 2008). We therefore do not have many empirical data points to help guide what we may expect from the Nmbo speech community. There are, however, some theoretical predictions. The speech of speakers in dense, multiplex social networks with frequent face-to-face interactions have been hypothesised to have particularly high levels of acoustic similarity across speakers (e.g. Bowern 2010, Trudgill 2011). Sociophonetic similarities may also be predicted for groups who place a high priority on solidarity, as has been said of New Guinean communities, e.g. Sankoff (1980a) observed that "[in Papua New Guinea] people pay very careful attention to small linguistic differences in differentiating themselves from their neighbours" (p.112). From this we may predict that speech communities like the Nmbo speech community will have high levels of acoustic similarity between speakers with very little variation at the community level. On the other hand, Dediu et al. (2013) have suggested that "features that mark group membership [such as accents] might be expected to be less strongly expressed in small communities" (pp.310-311), because there may be higher levels of individual variability due to familiarities with individual idiosyncrasies (see also Dorian 1994:283 on lack of social evaluation in small Gaelic-speaking communities).

Sociophonetic studies have been carried out on multiple levels of phonological organisation (segmental, suprasegmental, subsegmental, c.f. Foulkes and Docherty 2006, Thomas 2015 for consonants), and as the language communities under

investigation for sociophonetics expands, features such as tone are increasingly found to be loci of socioinguistic difference (e.g. Yang, Stanford, and Yang (2019) on Yangliu Lalo (Tibeto-Burman), Miller (2013) for Dane-Zaa (Athabaskan), Stanford (2007) for Sui (Kra-Dai)). Without a doubt, however, one of the most well studied is that of vowel quality. The decision to investigate the vowels of Nmbo was made for this reason. The same phoneme within a language variety shows fine-grained acoustic differences affected by social variables (e.g. Escudero, Boersma, Rauber, and Bion 2009 for Brazillian vs. Iberian Portugese, Adank, van Hout, and Smits (2004), and Adank, van Hout, and van de Velde (2007) for varieties of Dutch). Variationist vowel space studies have also been undertaken as a way of investigating social dimensions such as social diffusion in change phenomena (e.g. Northern Cities Vowel Shift (Labov 1994, Labov, Ash, and Bober 2006), California Vowel Shift (Eckert 2008), New Zealand English Short Front Vowel Shift (Maclagan and Hay 2007, Gordon et al. 2004), Montreal French (Yaegr-Dror 1996)).

Duration was included as a variable for investigation since a major distinction between the vowel classes of short and full vowel in Nmbo is based on duration. Vowel duration is known to have community specific socially conditioned variation, so the rationale is that Nmbo vowels may be one such language that exhibit some differences. For example varieties of English are known to manifest regional differences of vowel durations; British varieties of English show more variability in vowel durations compared to New Zealand English (Warren 1998), or Singaporean English (Deterding 2001). Certain varieties of American English show greater variability in duration than others (Fridland, Kendall, and Farrington 2014, Jacewicz, Fox, and Salmons 2011, Clopper and Smiljanic 2015). Gender is often found to affect vowel duration. There is a cross-linguistically well-attested observation that female speakers tend to have longer vowel durations than male speakers, with evidence from varieties of American English (Hillenbrand, Getty, Clark, & Wheeler, 1995), Creek (Johnson & Martin, 2001), Jamaican-Creole and Jamaican English (Wassink, 1999), Québécois French (Martin, 1998) and Swedish (Simpson & Ericsdotter, 2003). Some other studies have, however, found the reverse with male speakers having longer vowel durations, e.g. Northern British English (Whiteside, 1996).

The social variables that are known to manifest quite consistently across speech communities in vowel acoustics are gender, geographical region, and social categories of other kinds such as social class. This Nmbo sociophonetic investigation

will test whether gender, village of origin, village of residence, and speaker age affect the realisation of vowels in terms of height (F1) and fronting (F2), duration and vowel space spread where relevant.

6.2 Data: Additional Notes

The data used for this study are from the Nmbo Wordlist Corpus (NWC). The details of data collection and initial processing are outlined in section 5.2 and will not be repeated here. In this short section I will present the methods employed after the data collection phase.

The NWC includes a wide variety of words, but only a subset were chosen for the acoustic investigation. Some words, such as initial [h]-words showed a high degree of variability in vowel formant measures due to the variable [h]-drop, and are not included in this study. Table 5.2 in the methodology chapter 5 shows the words that were chosen for this investigation.

A total of 63 speakers were included in the final dataset. The speaker set constitutes 33 speaker from Bevdvn (15 female, 18 male), and 30 from Govav (12 female, 18 male). The years of birth (YoB) span 53 years, from 1943 to 1994. Table 6.1 presents speaker information. When organised by VoR, there are 11 women residing in Bevdvn, 10 in Govav, and 6 in Bimadbn (the Nen language village). The precise number of vowel tokens contributed by each speaker per word is presented in the appendix (A.10).

Vowel onset and offset were demarcated by the following criteria:

- Glottalisation was not included in the onset.
- Onset was marked after any release from the preceding consonant.
- Where vowel signal gradually trails off (i.e. word finally), offset of vocal fold vibrations were used to determine the vowel offset mark.
- In some cases intensity was used to determine the boundaries of onsets and offsets.

The vowel formants were normalised using Lobanov's z-score transformation. It is a vowel extrinsic, speaker intrinsic transformation that is identified by Adank, Smits, and van Hout (2004) as the best in maintaining sociolinguistic difference while suppressing physiological differences. The Vowel package (Kendall,

Thomas, and Kendall 2018) was used for normalising the F1 and F2 values, and the formulae are presented in the appendix (A.8). The measures of F1 and F2 have not been rescaled to approximate Hz values, following the recommendations of Kendall et al. (2018:16). While this makes the interpretation of the results slightly trickier, this removes any unintended numerical effects the re-scaling transformation may introduce to the data. The results are, therefore, presented in Lobanov scaled units.

For duration calculations, speech rate was normalised using the formula presented in Wassink (2006) (p.2345). The mean duration of each vowel category ([i, e, æ, ɑ, o u, ə, ɐ]) is calculated for each speaker. A grand mean of duration for each speaker is then calculated across all vowel categories. The mathematical formula used for this calculation is presented in the appendix (A.8). The value produced from the Wassink calculation is not a metric measurement of duration (i.e milliseconds) but a ratio of sorts. Therefore for the rest of this chapter I will call this ratio *duration units* when presenting numerical data on duration.

6.3 Results

We will begin with a brief descriptive overview of the vowel space based on the data, before proceeding to finer grained analyses of individual vowels and relevant sociolinguistic variables. The measurements for formants and duration presented in this overview are based on all the words from the NWC, with a variety of phonetic environments represented. Figure 6.1 graphically represents the mean height and fronting of all eight Nmbo vowels of 63 speakers. A summary table with descriptive statistics for this data are presented in table 6.2. The Nmbo vowel space from this data looks much like what is reported in Kashima et al. 2016[1], forming a triangular shape in acoustic space with clear distinctions shown by the full vowels [i, e, æ, u, o, ɑ]. The two short vowels [ə, ɐ] are also distinct from each other in acoustic space. The low full vowels [æ, ɑ] show tight clustering, as visible through the low standard deviation values of F2 which measure fronting and backing ([æ] SD = .072, [ɑ] SD = .081). The high vowels [i, u] also show tight clustering (for F2: [i] SD = .107, [u] SD = .132). There are very few outlier speakers, with most speakers falling within the confidence interval of 95% for each vowel.

[1]Note that [ə] in Kashima et al. 2016 corresponds to [ɐ] in this dissertation. The [ə] in this dissertation was not included in Kashima et. al.

Bevdvn				Govav			
Female		Male		Female		Male	
Speaker	Yob	Speaker	Yob	Speaker	Yob	Speaker	Yob
sfbgKS	1953	smbbSZ	1948	sfggDS	1947	smggWK	1943
sfbbKS	1956	smbbYZ	1951	sfgbMM	1949	smmgWW	1943
sfabKK	1957	smbbTS	1953	sfggZS	1949	smggWG	1947
sfabWP	1957	mmbbJS	1962	sfgmFY	1955	smggZG	1949
sfbaWZ	1957	mmbbLS	1967	mfgmGwN	1965	smggGM	1954
sfbbDS	1957	mmbbES	1970	sfgdWS	1968	smggMG	1954
sfamRA	1960	mmbbST	1971	mfggMW	1969	mmggMN	1968
mfagWA	1970	mmbbTY	1973	yfggJW	1985	mmggWW	1973
mfgbAN	1970	mmbbBD	1975	yfggED	1987	mmggGN	1975
mfbmYS	1972	mmbbLM	1975	yfggGY	1989	mmggGS	1976
mfbmHY	1976	mmbbKS2	1977	yfgmBZ	1990	mmggDZ	1980
yfbbJY	1986	mmbbKS	1979	yfggMW	1992	mmggGZ	1984
yfabRM	1989	mmbbKS3	1979			ymggSY1	1985
yfbbAT	1989	mmbbSS	1980			mmggRG	1985
yfbbYL	1991	ymbbTT	1982			mmggRW	1989
		ymbbMY	1983			ymggSY2	1991
		ymbbMZ	1985			ymggBS	1994
		ymmbSL	1989			ymggGW	1994

Table 6.1: Speakers in the sociophonetic study by village of origin.

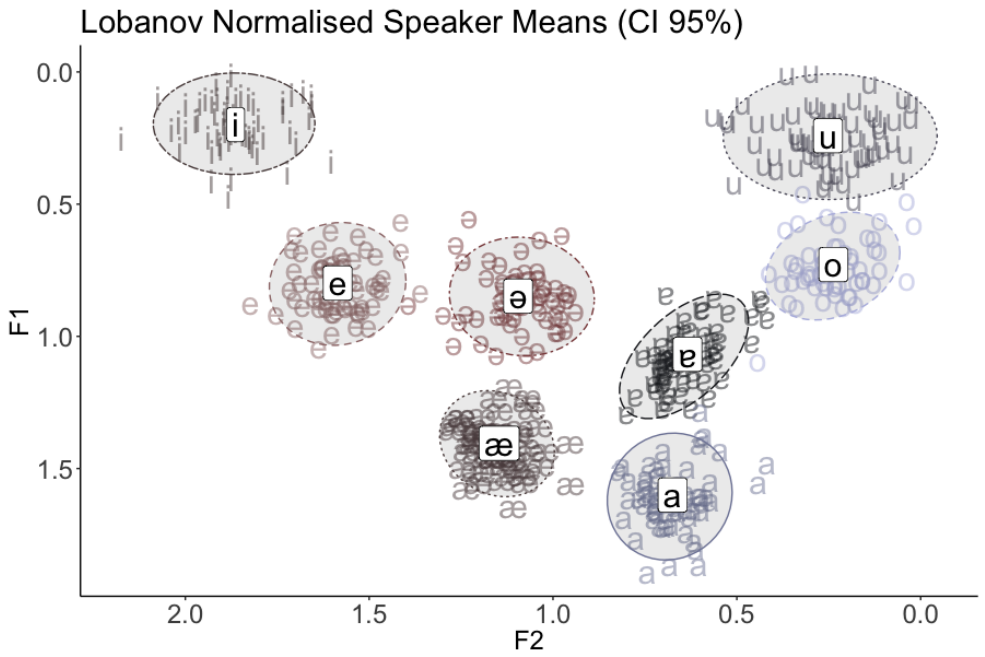

Figure 6.1: Speaker means per vowel for all speakers.

F1	n	mean	sd	median	min	max	range	se
i	63	0.199	0.092	0.194	0.011	0.468	0.457	0.012
e	63	0.798	0.104	0.808	0.559	1.035	0.476	0.013
æ	63	1.405	0.093	1.404	1.180	1.641	0.461	0.012
u	63	0.241	0.107	0.253	-0.006	0.480	0.486	0.013
o	63	0.728	0.106	0.740	0.450	1.088	0.638	0.013
ɑ	63	1.600	0.117	1.601	1.277	1.888	0.611	0.015
ə	63	0.848	0.108	0.852	0.546	1.060	0.515	0.014
ɐ	63	1.068	0.111	1.078	0.834	1.285	0.451	0.014
F2	n	mean	sd	median	min	max	range	se
i	63	1.863	0.107	1.875	1.603	2.174	0.572	0.014
e	63	1.586	0.087	1.596	1.361	1.823	0.462	0.011
æ	63	1.147	0.072	1.147	0.952	1.271	0.319	0.009
u	63	0.252	0.132	0.249	0.025	0.565	0.540	0.017
o	63	0.237	0.089	0.234	0.018	0.444	0.426	0.011
ɑ	63	0.675	0.081	0.692	0.418	0.821	0.402	0.010
ə	63	1.095	0.095	1.083	0.939	1.393	0.454	0.012
ɐ	63	0.634	0.083	0.653	0.418	0.800	0.381	0.010

Table 6.2: Lobanov normalised F1 and F2 mean values per vowel, aggregate of speaker means. F1 is the measure of vowel height, and F2 of vowel fronting. sd = standard deviation, se = standard error. The full list of means per vowel per speaker are available in the appendix (A.10, A.11)

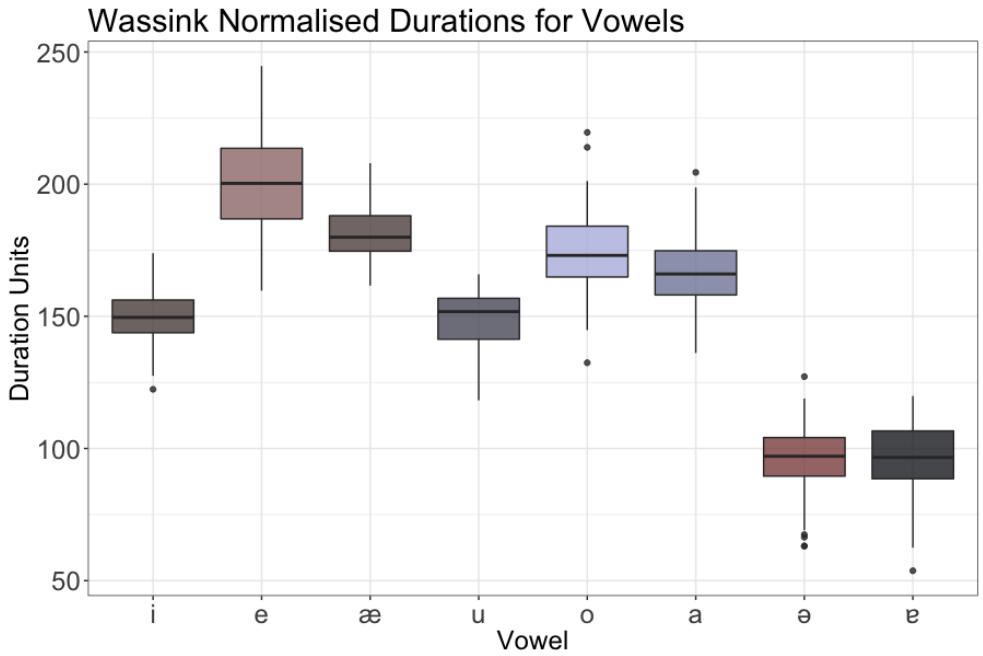

Figure 6.2: Speaker duration means per vowel for all speakers.

Vowel	n	mean	sd	median	min	max	range	se
i	63	149.08	42.48	145.17	34.03	302.08	268.05	1.84
e	63	201.23	57.82	182.68	108.35	364.51	256.16	3.31
æ	63	182.09	78.11	159.78	58.61	498.13	439.52	3.24
u	63	150.06	37.15	142.18	62.04	257.93	195.88	2.31
o	63	175.97	50.35	173.53	34.15	343.45	309.30	2.18
ɑ	63	164.36	48.02	152.75	67.82	347.90	280.08	2.16
ə	63	94.53	23.44	93.89	17.60	188.65	171.05	1.09
ɐ	63	96.77	32.51	93.27	3.72	230.66	226.94	1.10

Table 6.3: Descriptive statistics for normalised duration, aggregate of speaker means. sd = standard deviation, se = standard error.

Languages with large symmetric vowel inventories are hypothesised as having a higher realisation for a front vowel compared to a corresponding back vowel (e.g. Moraes, Callou, and Leite 1996, Seara 2000 for Brazilian Portuguese cited in Escudero et al. 2009). That is, the hypothesis predicts an [i] would be higher than [u] in acoustic space for a symmetric vowel inventory, [e] higher than [o], and [æ] higher than [ɑ]. The Nmbo results show these predictions as true for the [æ, ɑ] pair (F1 $\bar{x}$ = 1.405, 1.6), and the [i, u] pair (F1 $\bar{x}$= 0.199, 0.241). Visual inspection of the vowel space in figure 6.1 shows a clear height distinction between [æ, ɑ], and a slightly less clear but perceivable difference in [i, u]. The [e, o] pair do not clearly exhibit this tendency of the front vowel being higher in acoustic space. The mean F1 of [e] has a higher value than [o] ($\bar{x}$ = 0.798, 0.728), and visual inspection of the vowel space does not show a clear height difference. In other words the Nmbo mid-vowels [e, o] do not clearly fit the statement of a predicted tendency of slightly higher F1 values for front/back vowel pairs.

Figure 6.2 shows the distribution of durations for all vowels, and table 6.3 presents descriptive statistics of these durations. As expected, clear duration differences can be seen between the full and short vowel, with the short vowels on average around 56% of the length of full vowels. The average duration of the full vowels are 170.47 units long, while for the short vowels it is 95.65 units. The mean durations for both high vowels [i, u] are shorter than the mean duration of the full vowels ([i] $\bar{x}$= 149.08, [u] $\bar{x}$= 150.06). The remaining full vowels have quite different means from one another. A visual inspection of the boxplot 6.2 shows that [e] has a markedly different distribution to all the other full vowels. In the statistical modelling results we will find that, despite duration normalisation, monosyllabic words have longer durations of vowels. The longer duration of [e] is likely due to the over-representation of monosyllabic words in the dataset (*mer*, *end*, *men*, *nen*).

We will now look at the results of the linear mixed models. The goal of using these models is to quantitatively test whether there are patterns to the variability exhibited in the data (section 5.4). Initial model runs included phonetic parameters, such as preceding phonetic environment of the vowel, in order to determine whether there were any phonetic effects on vowel quality. These models did not explain the patterns of variability any better than models that did not include phonetic effects. The results presented from here on are thus based on models that do not include phonetic variables as fixed or random effects. The dependent variable is either F1 or F2 of non-rescaled Lobanov normalised vowels, meaning

the models are testing for what effects vowel height or fronting. The fixed effects are speaker Gender, Year of Birth (YoB), and Village of Origin (VoO) or Village of Residence (VoR). The random effects are Speaker and Word as random intercepts. The inclusion of Word is to try and capture lexical and/or phonetic environment effects in a general way.

The output of the statistical modelling for all vowels are presented in the appendix (A.13), but in the body of this chapter I will only refer to the findings of these outputs where relevant for the social variable under discussion.

The random effects of Speaker and Word explain the variability in the data to varying degrees. Some vowels and their measures have random intercepts predicted as highly significant (e.g. [e, o] height), while for others it was only marginally so (e.g. [æ] height, [o] fronting). Age and VoO are predicted as significant in a number of the models (summarised in table 6.4), but village of residence is never predicted as significant. In total, age and village of origin are predicted as significant for five different measures across four different vowels (e.g. [e] hight = 1 measure, [o] height and fronting = 2 measures). Gender is predicted for three different measures across three vowels. Overall, this suggests that age and village are relevant variables for explaining variation in the vowel space, but gender is less so.

While the results presented are statistically significant, my overall assessment of the vowel space is that it is stable with low levels of variation. This is due to the the low effect sizes of the various models, as well as the general fragility of the models. Since, however, the microvariations are of a systematic kind where the affected vowels can be groups into classes (e.g. high and mid vowels affected by age, short vowels affected by VoO), I take the results as reflective of some social reality of the Nmbo speech community.

6.3.0.1 Age

Age is predicted as significant for five measures across four vowels. It is predicted as a sole significant variable for [u] height (table 6.6), where the younger the speaker, the lower the realisation of [u]. For every increase in speaker year of birth, there is an increase of F1 by .002 units, resulting in a lowering of [u] by younger speakers. This decrease is statistically significant ($p < .05$), with the model explaining 63% of the variability in the data ($R^2c = .63$), although the fixed effects explain only 6% of the variability ($R^2m = .06$). This means that

Vowel	F1	F2
i	Model did not converge	Model did not converge
e	Younger** ↓ + random**	No significant fixed effects
æ	Govav** and Gen.* ↓ + random*	random*
u	Younger* ↓	No Effects
o	Younger** ↓ + random**	Male** and Younger* ← + random*
ɑ	Govav* ↑ + random*	Govav* → + random*
ə	Govav** ↓	Younger* ← + random**
ɐ	Govav ↓, Male** ↑	Unable to run model

Table 6.4: Summary of fixed effects predicted as significant per vowel, per measure. Arrows indicate the relative location in acoustic space of the vowel by the relevant variable, i.e. "Younger* ↓" means younger speakers have a lower realisation of the vowel relative to older speakers, "Govav →" means Govav speakers have a more backed realisation of the vowel relative to Bevdvn speakers. Random effects of Speaker and Word = 'random', i.e. "Younger** ↓ + random**" means speaker age and random effects were both predicted as significant . Significance codes: ** = p < .001 - .005 ; * = p < .01 - .05

	Random intercepts as non-significant	Random intercepts as significant	Total Times Significant
Age	u F1	e F1, o F1/2, əF2	5/16 measures, 4/8 vowels
Village of Origin	ɐF1	æF1, ɑF1/2 əF1	5/16 measures, 4/8 vowels
Gender		æF1, oF2, ɐF1	3/16 measures, 3/8 vowels

Table 6.5: Summary of vowel and its measure predicted as significant, per fixed effect. The first column shows the vowels and measures that were solely predicted as significant. The second column shows the vowels and measures where random effects were also significant. The third column tallies the number of measures where the social variable is significant.

there is a predicted .106 unit difference between speakers born across 1943 and 1994. More concretely, the entire range of [u] height is .486 units, so a lowering of .106 units across 53 years of age is a shift downwards that spans about 23% of the current height range. Figure 6.3 graphically represents the lowering of [u] by speaker age.

Age similarly effects the height of the vowels [e, o]. Unlike [u] where age is the sole variable for predicting lowering, [e] shows Speaker and Word as significant effects (for [e], $p < .001$, coefficient = 4.83, $R^2m = .05$, $R^2c = .51$; for [o], $p <$

Random	Intercept	Var	SD			
Speaker		0.008	0.093			
Word		0.002	0.044			
Residual		0.007	0.083			
Fixed	**Coeff.**	**SE**	**df**	**t**	**p**	
(Intercept)	-3.228	1.690	61.752	-1.910	0.060	
YoB	0.002	0.001	61.725	2.046	0.045	*
OriV:Govav	-0.0350	0.030	63.101	-1.348	0.182	
Gender:M	0.018	0.0262	2.758	0.684	0.496	

Table 6.6: Model fit summary for [u] F1 with fixed effects of YoB, VoO, and Gender. Speaker and Word as random intercepts. Speaker = 63, Word = 5, n = 258. Significance codes: * = p <.01 - 0.5. R^2m = .06, R^2c = .63

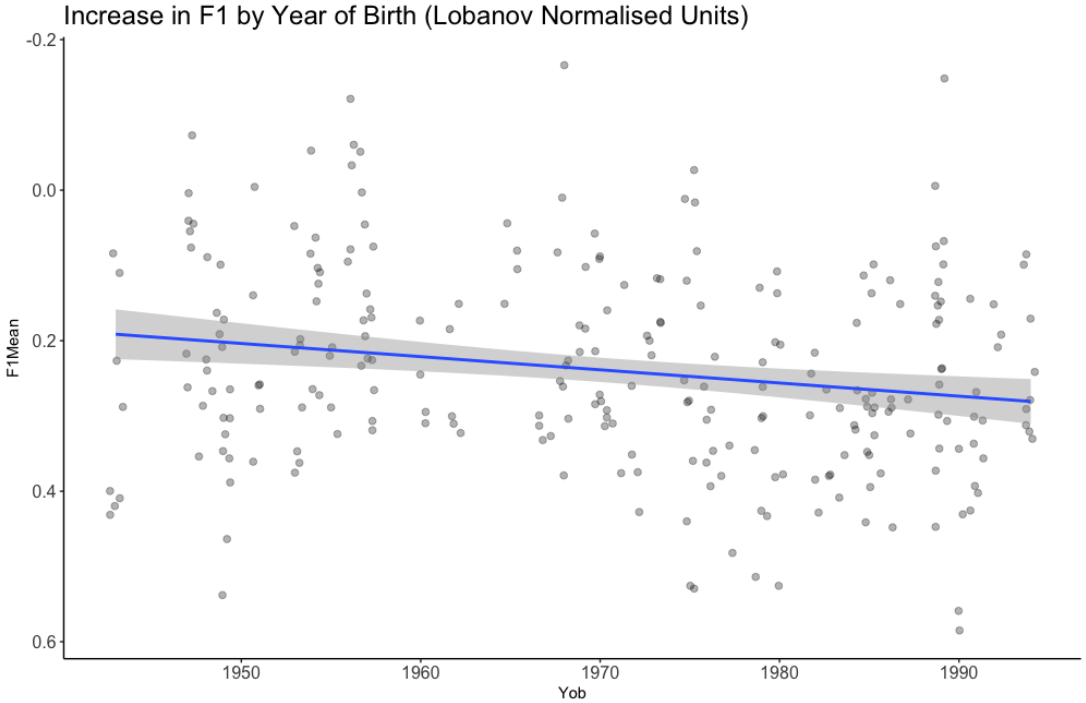

Figure 6.3: Decrease in [u] F1 by YoB. The y-axis has been inverted to reflect the vowel space, with an decrease in F1 value resulting in a lowering of the vowel in acoustic space. Each dot represents a mean per word per speaker. The line is the linear regression fit to the data with a CI of 95%. A Pearson's correlation of .2.

.001, coefficient = 5.1). Both [e] and [o] are predicted to increase F1 by .003 units per year of birth, resulting in younger speakers with lower realisations of both vowels (for both vowels $p < .001$). Again for comparison, the total range in height for [e] is .476, and [o] is .638, and an increase of .003 units per year results in a shift of .159 units downwards over 53 years. This is a larger shift than that predicted for [u] height. The full output of the model is presented in appendix A.13, table A.13 (page 429) and A.17 (page 431).

The two mid- and near-/back vowels [o, ə] show some evidence of fronting by younger speakers. Both have Speaker and Word as significant effects in explaining the pattern of variability. Younger speakers are fronting [o] more, with a .001 unit decrease in F2 value per year ($p < .02$, $R^2m = .06$, $R^2c = .46$)(table A.18, page 431). This results in a .053 unit decrease over 53 years, about 12.4% of the current F2 range (= .426 units). The short vowel [ə] shows a larger decrease in F2 values with .002 units per YoB, resulting in a fronting of [ə] by younger speakers ($p < .03$, $R^2m = .02$, $R^2c = .42$)(table A.22, page 433), about 23.3% of the current [ə] range (= .454 units).

Overall it appears that the high and mid vowels (excluding [i] which could not be modelled) are systematically showing an age difference. The full vowels [u, e, o] are lower for younger speakers, while [o, ə] show some evidence of fronting, also by younger speakers. Whether these Nmbo results are a case of age gradienting or a change in progress is an open question with the current data.

6.3.1 Village of Origin

Village of origin, like age, is predicted as significant for five measures across four vowels. Both short vowels [ə, ɐ] are predicted to show a village difference, where Govav speakers systematically have the lower realisation.

The model for the short vowel [ə] predicts VoO as the sole significant variable affecting height ($p < .0001$, $R^2m = .04$, $R^2c = .26$)(table A.21, page 433). Govav speakers on average have a .078 units greater F1 value than Bevdvn speakers, resulting in a lower realisation of [ə] than Bevdvn speakers. For [ɐ], we find VoO and Gender as both significant variables ($p < .02$ for VoO, $p < .005$ for Gender, $R^2m = .02$, $R^2c = .32$)(table A.23, page 434), and Govav speakers have a .05 units greater F1 than Bevdvn speakers, meaning, again, a lower realisation of [ɐ]. The data are visualised in figure 6.4.

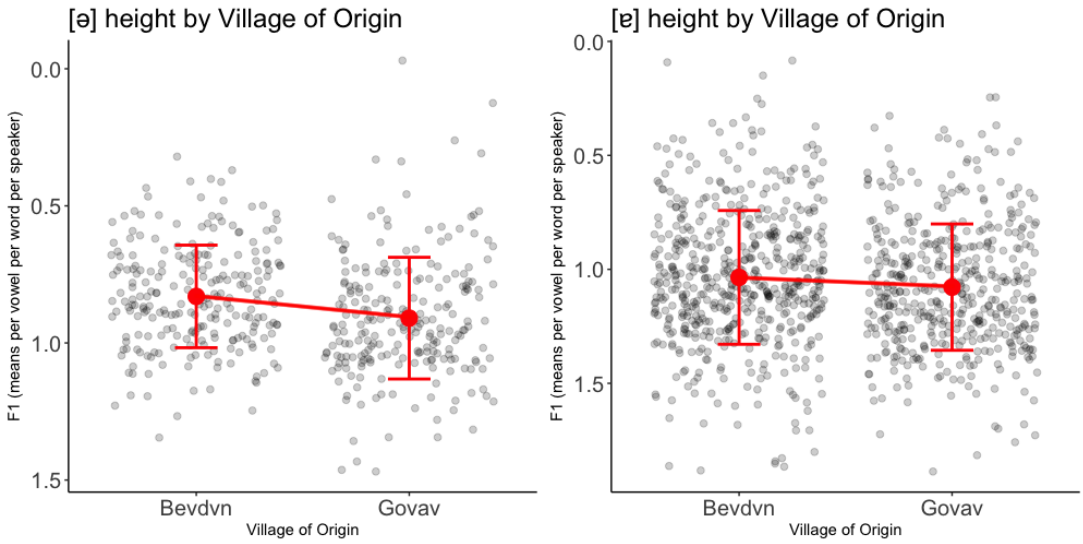

Figure 6.4: Difference in short vowels [ə, ɐ] F1 by village of origin. Each dot representing F1 means per vowel per word per speaker. The y-axis has been inverted to reflect the vowel space, with an increase in F1 value resulting in a lowering of the vowel in acoustic space. Error bar showing SDs.

The full vowels and their measures where VoO is predicted as significant are [æ] height, and [ɑ] height *and* fronting (tables A.15, A.19, A.20). Note that both are low vowels of the Nmbo vowel space. The Govav speaker coefficient of .063 units results in a lower realisation of [æ] than Bevdvn speakers ($p < .003$). The results for [ɑ] F1 are less robust ($p < .03$, $R2m = .03$, $R2c = 0.39$)(table A.15, page 430) but also show Govav speaker with lower average realisations in acoustic space by .046 units, resulting in a slightly backed realisation than Bevdvn speakers ($p < .03$).

The tendency for Govav speakers to have lower [æ], and more raised and backed [ɑ] contributes to a difference in vowel space spread across the speakers of Bevdvn and Govav. Vowel space spread for Govav speakers is less ($\bar{x}$= 105.63, SD = 58.65) than Bevdvn speakers ($\bar{x}$= 136.77, SD = 64.44); that is, on average Bevdvn speakers have larger vowel space spread. A one-way Kruskal-Wallis test[2] found a statistical significance between the two villages ($H(1) = 5.78$, $p < .02$).

	n	mean	sd	median	min	max	range	se
Bevdvn	32	136.77	64.44	106.02	59.30	267.11	207.81	11.39
Govav	31	105.63	58.65	80.96	43.07	226.82	183.75	10.53

Table 6.7: Descriptive statistics of vowel space spread by village of origin.

[2]An analysis of variance for non-normally distributed data.

Models with VoR were run to see whether locations of residence may explain the patterns of variation in the data. The linear mixed models were run only on female speakers,[3] since all but one of the male speakers in the sample reside in their village of origin. VoR did not come up as significant in any of the model runs, except it was nearly significant for [ɐ] F1 (table A.26, page 436).

6.3.2 Gender

Gender is a significant effect for three vowels and three measures. In all three models, Speaker and Word are both significant effects in explaining the pattern of variability. Unlike the results for age and village of origin, it is harder to find a systematic pattern to the differences across the two genders.

Gender is significant for the F1 of short vowel [ɐ], along with VoO (table A.23, page 434). The coefficient value of -.058 for male speakers shows that, on average, male speakers have a lower F1 value resulting in a raising of [ɐ] ($p < .005$, $R^2m = .02$, $R^2c = .27$).

Since village of origin was also predicted as significant in this [ɐ] model, the effect of gender was investigated separately by village. When we split the data into two sets, one of Bevdvn speakers and the other of Govav speakers, the effect of gender is different. For the Bevdvn sample population (n = 511, 32 speakers) there is a highly significant effect of gender ($p < .0001$, $R^2m = .05$, $R^2c = .26$)(table A.25, page 436), where male speakers on average are -.135 units lower than female speakers. The result is that male Bevdvn speakers show a slightly higher realisation of [ɐ] compared to female Bevdvn speakers. This effect of gender is, however, absent in the Govav sample population (n = 423, 31 speakers, $p < .039$). The difference between villages is visualised in figure 6.5.

A somewhat similar result emerge for [o] fronting where gender is predicted as significant, along with age, Speaker, and Word. This time, however, it is Govav speakers who show a gender difference while Bevdvn speakers do not. Models testing [o] F2 were run separately on Bevdvn and Govav speakers,[4] and male Govav speakers show more fronting on average (.06 units) than female speakers ($p < .02$) (table A.24, page 435). Gender as an effect is absent in the Bevdvn sample ($p = .13$).

[3]The fixed effects are YoB and VoR, and the random intercepts are Speaker and Word.
[4]Gender and YoB as fixed effects. Speaker and Word were set as random intercepts.

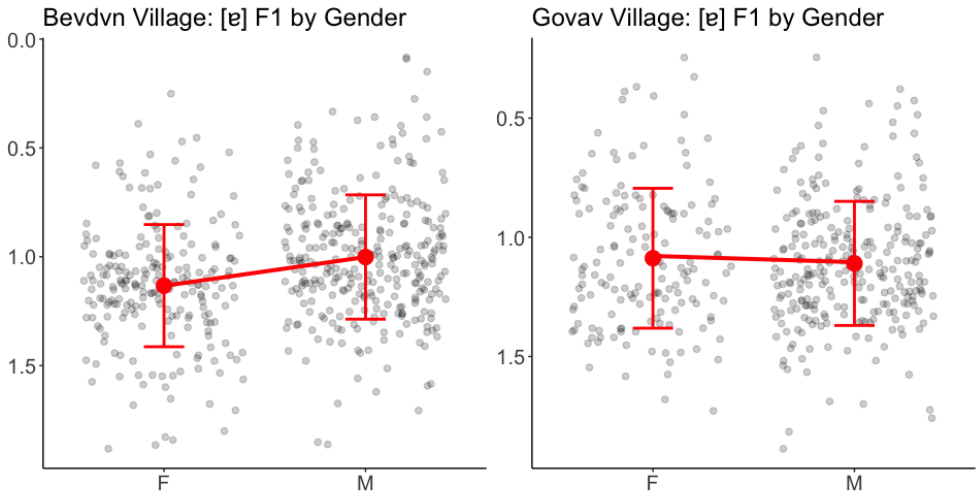

Figure 6.5: Difference in [ɐ] F1 by gender, by village. F = Female, M = Male. Each dot representing F1 means per vowel per word per speaker. The y-axis has been inverted to reflect the vowel space, with an increase in F1 value resulting in a lowering of the vowel in acoustic space. Error bar showing SDs.

	n	mean	sd	median	min	max	range	se
Female	27	109.08	61.45	84.91	43.07	267.11	224.04	11.83
Male	36	130.72	63.64	105.89	51.08	239.88	188.80	10.61

Table 6.8: Descriptive statistics of vowel space spread by gender.

Vowel space spread of female and male speakers were also compared. Cross-linguistic studies have found female speakers to have more open vowel spaces compared to male speakers of the same speech community, but the scale of this difference is known to be community specific (Henton 1985, 1992, Escudero et al. 2009). The vowel space spread for Nmbo female speakers has an average of 109.08 squared units (SD =61.45), while for male speakers it is an average of 130.72 squared units (SD = 63.64). That is, male speakers appear to have a more spread out vowel space, although a Kruskal-Wallis test did not find this difference statistically significant ($H(1) = 1.67$, p = 0.2). The story here seems to be that while there are small acoustic differences across the genders, these are minor and numerically non-significant.

A caveat here is that these results are based on calculations of Lobanov normalised values which may flatten out gender distinctions due to its function as a z-score calculation. Some sociophonetic studies use Bark normalised values to calculate dispersion instead, which are said to better reflect listener perceptions (e.g. Heffernan 2010:69 on phonetic and social characteristics of American ra-

dio DJs). It is possible that different normalisations, or different calculations of vowel dispersal would yield slightly different results.

6.3.3 Duration

Duration data are, unfortunately, much less robust than the formant data. The Nmbo duration data are heavily affected by lexical item and phonetic environment. Phonetic effects such as syllable number of word, stress, open or closed syllable were taken into account, and all of these variables have an affect on duration. In addition, all six full vowels and the short vowel [ə] had heteroskedastic residuals.[5] This means that statistical tests could not be used to interrogate the variability, and that there is some other pattern to the variation which I have not specified in the modelling. The short vowel [ɐ], however, could still be investigated with linear mixed modelling due to the normally distributed residuals. The results of the test again show Word as a random effect is significant. The model also shows phonetic variables as highly significant, and VoO as marginally significant.

The linear mixed model specified normalised vowel duration as the dependent variable. Unlike the dataset used for statistical modelling of F1 and F2, the words in this duration dataset are limited to those with obstruents preceding the [ɐ] (the words comprising this limited dataset is presented in the appendix, table A.8, page 418). The phonetic fixed effects included in the model were the interaction between Syllable (monosyllabic, disyllabic, or trisyllabic word), and Stress (stressed Y or unstressed N). The social fixed effects were VoO and Gender. Speaker and Word were included as random intercepts.

Unlike the formant results, we see a lot of variability in the duration data that the model cannot capture. This is evidenced by the high standard deviations of the residuals (table 6.9). In addition, the random effect of Word explains a lot of variability in the data with an incredibly high SD of 18.20. The plotting of the random effect of word (figure 6.6) shows *qɐv* ('hole') and *tɐnd* ('hand') showing significantly divergent durations of [ə] compared to the other words in the set.

[5] In statistical parlance, heteroskedastic residuals indicate that the data are non-normally distributed, and that there are some patterns to the data which the model was unable to capture. Inferential statistics such as regression models are interpretable if the data are not heteroskedastic, ie. homoskedastic. There are alternative models and methods to interrogate non-normally distributed data that are used by statisticians, which were outside the purview of this dissertation and timeline.

The word *qɐv* has [ɐ] preceded by a coarticulated consonant <q> [k͡p], but *qɐt* is also a phonetically quite similar word showing a duration distribution much closer to the mean. Looking at the words we can also see that the monosyllabic words all have longer durations than the mean, and we can see this results in the output of the fixed effects.[6]

Random	Var	SD				
Speaker	102.6	10.13				
Word	331.3	18.20				
Residuals	282.2	16.80				
Fixed	**Coeff.**	**SE**	**df**	**t**	**p**	
(Intercept)	129.690	7.061	22.090	18.368	0.000	**
Stress Y	-38.249	3.727	534.113	-10.264	0.000	**
SyllMono	13.462	3.426	509.141	3.929	0.000	**
SyllTri	-62.592	19.558	11.434	-3.200	0.008	**
OriVGovav	-7.574	2.904	59.870	-2.608	0.011	*
SyllTri:StressY	60.941	26.218	11.200	2.324	0.040	*
Sex	5.515	2.927	59.390	1.884	0.064	

Table 6.9: Model fit summary for [ɐ] duration with fixed effects of VoO, and Gender. Speaker and Word as random intercepts. Observations 653, Speaker, 63; Word, 12. Significance codes: ** = p < .001 - .01 ; * p < = .01 - .05

The random effects are highly significant ($p < .0001$), but the fixed effects of Stress and Syllable are also very strong. When the model tested Stress and Syllable separately, both were predicted as highly significant (table 6.9). Stressed vowels tend to be shorter by 38.25 units, and trisyllabic words shorter by 62.59 units. For reference, the total range of [ɐ] duration is 226.94 units (table 6.3) Mono syllabic words, on the other hand, tend to have a longer duration (figure 6.7). The results of Syllable show that, despite the normalisation of duration, the effect of word length still affected the results.

The result of this statistical modelling also suggests that longer duration is not a significant correlate for stress in Nmbo, at least for [ɐ]. Given that short vowels are able to take stress just like the long vowels (section 4.1.2.1), this is not

[6]Another approach, suggested by one examiner, would be to include another fixed factor comprised of a set of phonological features, such as coda sonorant versus obstruent, or coda voiced versus voiceless. The suggestion was that this may dampen the influence of word as a random effect and enable the fixed effects to contribute more meaningfully to the model results.

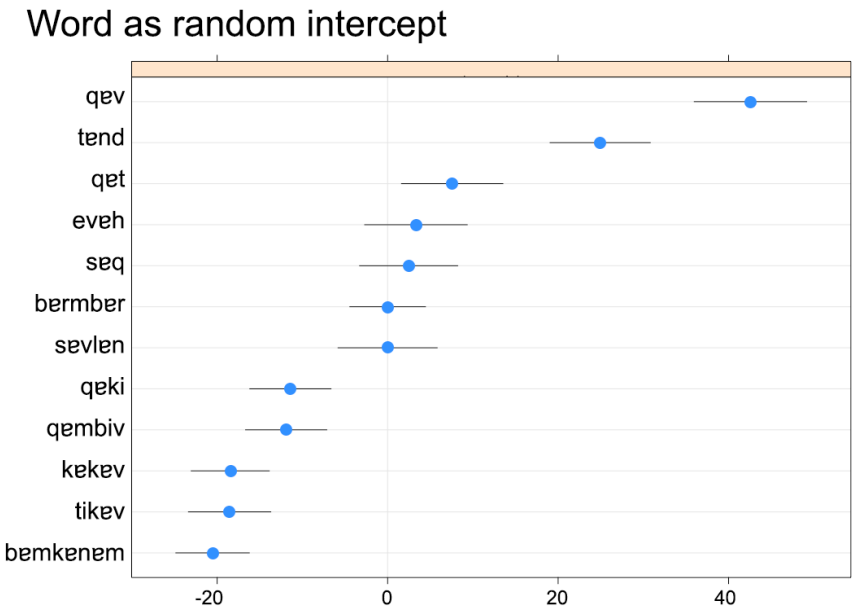

Figure 6.6: Conditional modes of the random effect of Word. The further a word is from the mean of zero, the more divergent the duration of that word is from the mean. Positive values mean longer durations, while negative values mean shorter. $R^2m = 0.30$, $R^2c = 0.73$

entirely surprising. Proposed cross-linguistic correlates of stress are pitch, intensity, and duration (Gussenhoven 2004: 14-15), but different languages manifest these correlated in different ways. For example longer duration and higher pitch are correlates of stress in Pitjantjatajra (Pama-Nyungan; Australia)(Tabain, Fletcher, and Butcher 2014), while correlates of Kuot stress include duration but not pitch((Papuan; PNG, Lindström and Remijsen 2005). Pitch and intensity were not investigated for this Nmbo study, but a preliminary hypothesis from this modelling is that longer duration is not a strong correlate of stress in Nmbo.

The statistical modelling also predicts VoO as an effect, with villagers from Govav showing a shorter duration (coeff. -7.57) compared to villagers from Bevdvn. Figure 6.7 show that these differences are small, but this result of VoO as significant follows the pattern of the vowel formant models. That is, village is a social variable which the models predicted as significant in explaining variation.

The overall strong effect of lexical item on duration suggests an unfortunately homogenising effect of the word list format. It is also possible that the preceding consonant is having some kind of effect. The strong effect of syllable count of word also suggests that despite duration normalisation, monosyllabic words

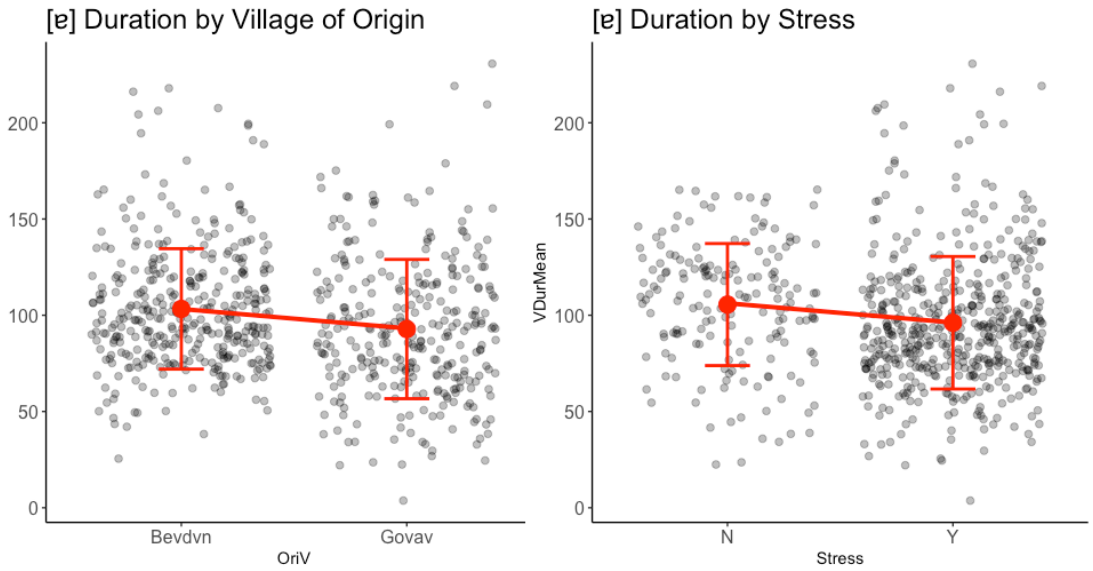

Figure 6.7: Difference in [ɐ] duration by village of origin, and stress. N = not stressed, Y = stressed. Each dot represents mean duration per vowel per word per speaker. Error bar showing SDs.

continue to have a lengthening effect on vowel duration. The other finding of interest is the effect of stress on vowel duration. This suggests that duration is not a strong correlate of stress in Nmbo. The implication of this finding is that other phonetic variables such as pitch or intensity may be more relevant in characterising Nmbo stress.[7]

6.4 Discussion and Conclusion

The overall picture to emerge is that the Nmbo vowel space is stable with evidence of socially conditioned micro-variations. The social variables of age, village of origin, and gender are all predicted as significant for different measures across different vowels. We cannot say anything about the effect of village of residence due to the fragility of modelling, but from this data it appears that VoR is not a significant variable.

When taking the YoB results for all relevant vowels as a whole, there appears to be a systematic difference in the vowel space between older and younger speakers. The full vowels [e, u, o] show lowering by 0.002-3 units per YoB, and [o, ə] show evidence of fronting by 0.002-3 units per YoB. It is not entirely clear, however,

[7] An examiner mentioned that preceding consonants have major effects on vowel duration, and with hindsight I notice the major error in not considering the preceding phonetic environment of vowels more thoroughly. I believe my results are so inconclusive in this particular section precisely because I did not account for preceding consonant.

that this is a change in progress. The *apparent time construct* (Labov 1972b, c.f. Cukor-Avila and Bailey 2013) may suggest that the downward shift of the high- and mid-vowels is a change in progress, since there is a systematic difference between the oldest and youngest speakers in the community. There is evidence that the apparent time construct is reliable for phenomena that show a clear *s*-curve distribution of new variants across age groups (Pope, Meyerhoff, and Ladd 2007, Blythe and Croft 2012b), but we cannot readily tell with this vowel data since the variations are, in absolute terms, very small.

An alternative way to interpret the differences based on speaker age is to consider age cohort effects. That is, older and younger speakers may be speaking in some community-defined age-appropriate manner. While not a phonetic-phonological example, Wagner and Sankoff's study of periphrastic vs inflected future in Québécois French showed that some speakers, as they aged, started adopting the more conservative inflected form, which went against the predicted trend of the rise in periphrastic forms (Wagner & Sankoff, 2011). Meek (2007), in discussing a language shift situation in the Yukon in Canada, demonstrates how young heritage speakers of Kaska saw the use of Kaska as tied with age and authority. The sentiment behind the statement "you start speaking Kaska as you get older," (Meek 2007:34) is that the usage of Kaska speech is limited to those in a position of cultural authority. As I mentioned briefly in the general discussion of age in the Nmbo speech community (section 3.3.1), age and authority go hand-in-hand, but there is the possibility that there are some locally pertinent age cohorts. While I did not explore those in this study, some possible candidates are "married vs unmarried", "speakers who received schooling during the Australian colonial occupation", or perhaps even "raised before or after the independence of New Guinea".

Another possibility is that the speech of younger speakers are less set than those of older speakers. Cukor-Avila's studies of changes in vowel realisations of AAVE speakers in a Texas community of 180 speakers suggests that older speakers have quite a bit of stability in their vowel realisations over real time, while teenage speakers show a lot more variation (2000, 2004). The younger speakers of Nmbo who are sampled here cannot be compared to teenagers in North America, but the concept of younger speakers having less-set realisations of speech is possibly applicable. Unfortunately the current available method to assess whether synchronic micro-variations conditioned by age are changes or not, is to collect data in the future and undertake a longitudinal study. We cannot make the call

for the Nmbo data here, but it is clear that there is some systematic age-based variation.

Village of origin is predicted five times as significant for explaining variation in formant measures. Of these five, three predict Govav speakers as realising [æ, ə, ɐ] lower than Bevdvn speakers. In other words there is a systematic difference between Govav and Bevdvn speakers in the locus of short vowels, where both [ə] and [ɐ] show a height difference. Models for [ɑ] predict a different direction, with the realisations by Govav speakers predicted as higher and backer than Bevdvn speakers. The vowel spread data suggests that Govav speakers have a smaller vowel space, making VoO the most frequently recurring social predictor of significance. The identification of differences between the villages do not tell us whether there is a directionality to these differences. We cannot say, for example, that Bevdvn speakers are spreading their vowel space from a stage that was more like Govav speaker, or that Govav speakers are lowering their [æ]. The numerical differences are indicating a difference in vowel realisation between the two sample populations.

The systematic difference in height of both short vowels provides some evidence that Bevdvn and Govav are indeed micro-speech communities within the greater speech community of Nmbo. The short vowels, and [ɐ] in particular, are the most high-frequency vowels that occur in Nmbo utterances[8]. The high frequency is due to the unique phonological status of [ɐ] as a predictable, but occasionally phonemic, vowel in Nmbo phonology (section 4.1.2.1). Inflected verbs all contain predictable short vowels, and other high frequency lexical items such as demonstratives also contain short vowels (e.g. distal demonstrative *kɐt*). The high levels of exposure of Nmbo speakers to [ə, ɐ] may make these vowels particularly susceptible to the effects of accommodation. Correspondingly the high usage of [ə, ɐ] would make these vowels susceptible to articulatory reinforcement of producing the community norm.

The effects of gender are not systematic across the data unlike YoB and VoO. Realisations of [ɐ] by male Bevdvn speakers are higher than female Bevdvn speakers. The models for [æ] for Bevdvn speakers suggest male speakers have lower realisations than female speakers. Realisations of [o] by male Govav speakers are more backed than female Govav speakers. Again we cannot say from this data

[8]As mentioned in the phonology section, the impression is that [ɐ] is the default form of the predictable vowel. In neighbouring Nen, the most common form of the short vowel is [ə] (Evans and Miller 2016: 19), which corresponds to Nmbo [ɐ].

if there is any directionality in this difference (i.e. male speakers are lowering vs female speakers are raising). Unlike YoB and VoO, there is not a single statistical model that predicts Gender alone as a predictor. For the models that predict Gender along with other variables as significant, the effect is lost once the sample population is split into two along village lines. Even if we take the models at face value, and accept that [ɐ] F1, [æ] F1 and [o] F2, have a gender difference, this is not particularly informative since there is no phonetic-phonological relationship between these three vowels and the formants that are affected. It is possible that these differences have some effect on the vowel space spread, where the lower [æ] and fronted [o] by men results in the larger vowel space found by calculating the area of a trapezoid. The results of vowel space spread and duration, however, also show no significant gender differences. This is not to say that gender is unimportant in the Nmbo speech community, but that gender differences do not manifest in the particular measure of vowels that were investigated.

The relative non-significance of gender is quite striking considering the robustness of gender as an effect in studies of English and other European languages. The absence of gender as an effect is rarely reported on in a clear way, but are attested in various speech communities. Hawai'i Creole (Pidgin) is reported to lack strong gender distinctions (Grama, 2015), and the GOOSE vowel in Black and White South African English are also reported as lacking gender distinctions (Mesthrie 2010: 18). The explanation offered by Grama regarding the lack of gender in Pijin vowel acoustics is that of cultural orientation (Hazen, 2002). Grama argues that Pidjin indexes social meanings such as "localness", which is more important in the language ecology where standard American English looms large as a presence (2015: 258-259). In other words, marking gender through linguistic means is secondary for speech communities that have larger targets to culturally orient against. Indeed, the recurrence of VoO as a variable in the Nmbo data supports the notion that geographical place is more important in Nmbo internal variation. Orienting against the other gender within the Nmbo speech community does not appear to be a linguistic priority. Orienting against the Other Village, on the other hand, *is* a priority. We will see this pattern again in the following [h]-drop study, so the non-significance of gender in relation to cultural orientation will be discussed again in the conclusion (chapter 9).

The lack of village of residence as a significant result is likely partially due to the word list format of the data. Speakers are highly monitoring their own speech, and are more likely to produce forms that match community norms . Theories

of second dialect acquisition (Chambers 1992, cf. Nycz 2015, Siegel 2010) may predict that the Nmbo speaking women residing in Bimadbn will shift their articulation of short vowels towards those of Nen speakers, but we do not see evidence of this from the data. The full vowels of both Nmbo and Nen are very similar, but the short vowels are known to vary across the two languages. Descriptions of the Nen vowel space represent the short vowels as [ɪ] and [ə] (Evans & Miller, 2016), which correspond to the Nmbo [ə] and [ɐ] respectively. That is, the Nen short vowels are higher and fronter than Nmbo short vowels in acoustic space. A preliminary study of Nen vowel acoustics support this description (Ellison, Evans, Kashima, Schokkin, & Williams, 2016). The results from this Nmbo study suggest that these daily bilingual women pronounce their [ɐ]s very much like the Nmbo speaking women residing in Nmbo villages. Natural speech data will need to be processed in order to gauge what the daily bilingual speakers are actually doing in the naturalistic speech of both languages. In the meanwhile, this study shows that in careful speech, the short vowel [ɐ] of the daily bilingual women are well within the norm of their peers residing in Nmbo villages.

My overall impression is that the vowel space is a stable part of the Nmbo linguistic system. There are a couple of possible reasons and interpretations for this stability. One possibility is that small vowel inventories are more conducive to stability, but given the absence of studies that have systematically investigated the relationship between variation and vowel inventory size, this possibility is only floated as a hypothesis for future investigations. It is not entirely clear that the Nmbo vowel inventory size of eight is small, given an inventory of this size tends to be classified as large from a cross-linguistic point of view. For example Maddieson (2013) classifies a vowel inventory size of 7 to 14 as "large" for the purposes of WALS groupings, which is the group Nmbo would belong to. Another possible reason for the stability of the vowel space is in line with Bowern (2010) and Trudgill's (2011) hypotheses that small-scale societies show high levels of linguistic similarity. The significant results of Speaker as a variable for many of the vowels, however, suggests there is much intra-speaker variation, which aligns with Dediu et. al's (2013) hypothesis that there are high levels of idiosyncrasies in small communities. The Nmbo results can be interpreted either way, and a more sophisticated investigations into individual speakers is necessary to provide more robust support for either hypotheses.

Having said that the vowel space appears stable, I do not believe that Nmbo

speakers have very similar phonetics and pronunciations overall. If I were to deliberately seek highly variable phonetic features, I would choose to investigate Nmbo consonants, especially the class of fricatives (A.19). The fact that the sister languages of Nmbo have very different consonants that correspond to Nmbo fricatives (4.1.1.1, 4.1.1.4) suggest that these are susceptible to variation for a variety of reasons. The point to emphasise from this study is that the vowel space is not an area of high phonetic variability, but the micro-variations we can detect are still informative of the social factors at play within this small speech community.

This vowel sociophonetic study is the first conducted on a Papua New Guinean speech community that speaks a vital vernacular language. Speaker and word are major effects, but age, and village of origin, are also predicted as significant. The mid-lower portion of the vowel space is realised lower by younger speakers, and Govav speakers appear to have lower realisations of the short vowels [ə, ɐ]. Despite there being some gender differences when the sample population is modelled as a whole, the gender effects are prone to disappear when populations from each village are investigated separately. The conclusion is that gender differences are not significantly manifested in Nmbo vowels, and we will see in the next study again that gender is not detected as a significant variable in explaining the patterns of variability.

Chapter 7

Study 2: Word Initial H-Dropping

Gnoso yam gym, bä fonatae naäyätusa. Ämb ḡä v ym. "The way of things today, it has become different. It seems so." (Maiwa Yaki, Multilingual Questionnaire 2017)

In this chapter I explore the sociolinguistic conditioning factors around word initial [h]-drop in lexical nouns. Unlike the vowel sociophonetics, which show a stable vowel system, [h]-drop is a change in progress phenomenon that is found throughout the Nmbo speech community. The aim is to provide a detailed account of the distribution of linguistic and sociolinguistic features associated with this variable, and to discuss their significance in relation to our understanding of socio-cultural selection pressures affecting language change in the Nmbo speech community. In other words, the goal of this chapter is to identify which speakers are at the forefront of [h]-drop, as well as to describe any linguistic factors that may be affecting differential rates of [h]-drop.

The variable under investigation is word initial [h]-drop variability. Nmbo speakers exhibit variation in their realisations of lexical nouns beginning with [h], where some utterances are produced with the [h] while in others it is dropped.

(178) Speaker DS is undertaking the *Cassowary Picture Task*, and is describing the picture she sees in front of her. Parenthesis indicates instance of dropped [h]. Note that both [h]-words occur intonation unit (IU) initially.

a. **hrare** site-yan ädi y-ng/m bei.
moon.ABS light=LOC EMPH 3sg.nphd/go DISC
In the moon light, she is going [i.e. walking, travelling].

WSEK1-G20170707-05Cassowaryhm, 00:07:47.359 - 00:07:48.895

b. **(h)rare** yna site-yan ädi y/uta
moon.ABS PROX.DEM light=LOC EMPH 3sg.nphd/walk
bei.
DISC
In this moon light she is walking.

WSEK1-G20170707-05Cassowaryhm, 00:07:49.455 - 00:07:51.935

The rationale section will further articulate the reasons for choosing [h]-drop as a variable, and contextualise this study in relation to the few systematic studies of [h]-drop available (section 7.1). Results from a word list study are presented, which show that the highly structured elicitation format still resulted in variable rates of [h]-drop across the sample population, supporting the impression that there is no social commentary or meaning associated to this phenomenon. The results also suggest strongly that [h]-drop is a change in progress below the levels of consciousness, rather than a case of stable variation (7.2). Before moving on to the results of the naturalistic speech study, a short methodology supplement to the details of the Nmbo Sociolinguistic Corpus (NSC, 5.3) will be presented, with attention to the annotation of [h]-drop specifically (7.3). Results are in two parts. The first part explores the linguistic environments in which [h]-drop occurs (7.3.1). The second part concerns social factors (7.3.2). The results of the naturalistic speech data support those of the word list study, where age is the strongest variable related to [h]-drop propensity. The discussion and conclusion (7.4) will explore the possibility of Govav as the origin of [h]-drop, with a closing comment on tempo of change.

The relevant sections of the sketch grammar for this study are the phonology section (4.1), in particular the consonant section (4.1.1), and the word class section (4.2).

7.1 Rationale and Context

Nmbo [h] is a phoneme found in lexical words such as nouns and verb roots. It can occur word initially, intervocalically, and word finally. Nmbo [h] likely derived from Proto-Nambu *s (Evans et al., 2017), and corresponds to the velar approximant /ɯ/ in Yarne Nambo/Namna, and the voiced velar fricative /ɣ/ in Nama, and /s/ in Nen. The [h] occurs in other Nambu languages, but in limited contexts such as affirmations and loan words (e.g. in Nen, Evans 2015b). Nmbo [h] should be an emblematic phoneme in some sense, and yet Nmbo speakers are dropping it at various rates in their speech.

Nmbo speakers themselves do not remark upon the non-realisation of word initial [h]. The earliest social evaluation of [h]-drop I heard came from a licensed Nen speaker, Jimmi Nébni, who comment to me in the first year of my fieldwork, "Why are those Kerake dropping their h?". This anecdotal episode suggests two things. One, that the direction of change is indeed of [h]-drop rather than [h]-gain. Second, it shows how non-Nmbo speakers might view [h] as a marker of Nmbo, and are sensitive to its realisation in a way the Nmbo in-group are possibly not. From my observations in the field it is not entirely clear whether the Kerake themselves view /h/ as emblematic of their language. As Dorian (2010) notes, speakers in small and dense-knit speech communities are often more sensitive to variation in geographical terms (inter-village variation), but less so to local variation in individual terms (intra-village variation)(p.33). Dorian's statement applies to the Kerake villages and the Morehead area more generally, where social evaluation and identification often focusses across varieties associated with different tribes or villages, rather than what speakers in ones own tribe or village are doing.

The lack of social evaluation of [h]-drop also suggests that the phenomenon has proceeded through the speech community; that is, it is a change in progress. A classic observation made by Labov (1965) is that a "change from below" (i.e. below the level of social awareness) shows no patterns of stylistic variation (p.110), and that such change proceeds across time by the "incrementation by children, who reproduce and advance their parents' system" (Labov 2007: 379). Change

from below can thus be characterised as *variation conditioned by age*, but where there is no social meaning associated with variable realisations. Labov has speculated elsewhere that "imitation and reference group association" may be less prominent in change involving lenition and deletion specifically (Labov 2001:26-7), so the Nmbo [h] may be an instance of this. Lenition and deletion phenomena can, however, generate social meaning in the long term. Dialect differences and cognate sets of phonemes are often based on results of past lenition changes, as is clearly the case with the Nambu languages. The ascribing of social meaning occurs after the fact when geographical location of speakers matches resultant change, rather than during the linguistic change process which may not be accompanying any other social fissuring of a speech community. What is interesting in Nmbo, then, is whether there are other latent social characteristics of [h]-drop that hint at the origins of the phenomenon.

I have chosen to limit the investigation to word-initial environment of [h] for practical reasons. Word final [h] is hard to detect, and relatively infrequent in Nmbo speech. Intervocalic realisations of /h/ can occur in lexical nouns and infinitive verb stems (e.g. *on-h=ɐt* 'to gather' suffixed with the allative {=ɐt} to produce 'to go gathering', section 4.4.1), but these constructions are somewhat infrequent in natural speech. Furthermore, there is evidence that even "the same phoneme" may undergo sound changes in different ways depending on the position of the phoneme in a word (e.g. Greek vs Spanish type pathways of /s/ to /h/ lenition, Ferguson 1990, syllable final vs word final /s/ in Spanish varieties Brown and Brown 2012). There is no reason to assume that intervocalic Nmbo [h] would change under the same pressures as word initial [h]. This Nmbo study aims to limit the envelope of variation so that more can be said with certainty within its scope.

Sociolinguistic studies of [h]-drop have come almost exclusively from studies on English. The presence or absence of [h] has been described as one of "the foremost signals of social identity" (Mugglestone 2003:95) in British Isles Englishes, with the dropping of [h] being a strong marker of regional varieties or class membership. This finding for English quite clearly cannot be generalised to other speech communities, and there do not appear to be any other systematic studies of [h]-drop in languages other than English. Quantitative studies of [h]-drop exist in a number of British English varieties (e.g. Trudgill 1974 for Norwich English, Coupland 1988 on Cardiff English, Milroy 1992 for overview of [h] in variety of Englishes) or in other English varieties (e.g. Bell and Holmes 1992 for New

Zealand English, Horvath 1985 on Australian English, Irvine 2004 on Jamaican English, Buchstaller 2018 on Marshallese English, Lynch 2016 on Kosrean English). There has even been a historical corpus study of [h]-loss in Middle English (Crisma, 2007). Wells (1982) observed that [h] deletion is influenced more by social factors rather than linguistic factors, so naturally the studies on English-varieties focus on social meanings attached to variable [h]. This tendency seems to have resulted in a paucity of variationist studies with variable rule analysis of [h], since realisation of [h] in English varieties rarely vary. The understanding is that Englishes are either [h]-ful or [h]-less.

Nonetheless there are a handful of variationist studies that look at phonological conditioning factors, and we can make at least one hypothesis from these. Bell and Holmes (1992) report that preceding liquids, obstruents, and nasals, favour [h]-drop in New Zealand English. Buchstaller (2018) reports that preceding nasals as slightly favouring [h]-drop in Marshallese English. Interestingly, Trudgill (2004b) suggests that the loss of word initial [h] in Old English occurred due to following /r, l, n/ environments (p.73). It is possible that preceding nasals are a common and relevant linguistic conditioning factor in [h]-drop due to the articulatory characteristics of nasals.

What is known about the phonetics of [h] indicate that it is highly prone to dropping or elision in many environments. There is some disagreement over the analysis of [h] as a voiceless abutting vowel (Ladefoged 1982), or as a voiceless fricative (c.f. Laufer 1991), but both characterisations agree that [h] production involves abducted (i.e. spread) vocal folds, and that voicing is either absent or highly reduced (Pierrehumbert and Talkin 1991:93-94). Effort based accounts of phone realisation (Kirchner, 2004) would therefore assume that any phonetic production requiring a shift between a voiceless or open glottis state might find a speaker dropping or eliding the [h] (e.g. an open vowel [h] open vowel). Returning to the aforementioned observation of nasals favouring [h]-drop, nasals and [h] differ greatly in place and manner of articulation, so ease of articulation would suggest that speakers reduce glottal abduction, if not completely skip it. Another phonetic explanation for what favours [h]-drop is that a preceding lengthened vowel may absorb [h] into the vowel (e.g. Ferguson 1990:63). Pierrehumbert and Talkin (1991) support this evaluation generally, but suggest that suprasegmental factors such as accent and intonation boundaries need to be considered more seriously since these suprasegmental parameters have a large effect on the individual productions of [h]. In sum, the acoustic variability and weakness of

[h] is undoubtedly what makes [h]-loss a common phonetic phenomenon.

7.2 Word List Study Results

A word list study was conducted as an analogue to reading passages in Labovian sociolinguistic interviews. A subset of the Nmbo Word list Corpus (NWC) was used for this end. Reading passages often reveal some kind of shared sociolinguistic norms held by the community whether it be a propensity for some speakers to produce standards and prestige varieties(Labov 1966, 1972), or as an opportunity for speakers to reject prescribed standards and prestige varieties by actively producing local vernaculars (Milroy 1987 for English in Belfast,Gafter 2016 for Hebrew dialects). Shared sociolinguistic norms are of interest for researching variation in a language without prior documentation, and this study provides an example of what highly monitored speech looks like from a speech community with no traditional orthography or education system. In this particular case, this word list study also results in providing a preview for the findings of the natural speech data.

The word list is an elicitation situation (section 5.2 for details), and corresponds to reading passages conducted in sociolinguistic studies of literate communities. The reading passage and word list are both situations where speakers are highly conscious of their speech production, and a task where perceived correct forms and norms are likely to be produced. Both formats have some representation of the target form modelled; the orthography in reading passages, and the verbal administration by the linguists in the word list. Five [h]-words are covered in the pilot: *hakr* 'boy', *hamba* 'village', *hkv* 'eye', *hrare* 'moon', and *hure* 'true', 'straight'. These five words were identified as having word initial [h] by comparing it against cognate words in Nen, which all have word initial /s/: respectively, *sakr*, *samba*, *skop*, *srare*, and *suri*. Each speaker produced three repetitions of the word without a carrier phrase, producing a total of fifteen tokens. This data was collected between the years 2014 and 2015, and consist of 50 speakers from across the villages of Govav and Bevdvn, born between 1941 and 1994 (aged between 21 to 74 in the year of recording).

Figure 7.1 plots speaker by number of [h]-ful words. Each point represents a speaker, and the x-axis indicates the year of birth. The y-axis has been inverted to be visually similar to the natural speech data results in the next section, so the lower value of [h]-ful words means a higher propensity of [h]-drop. A numerical

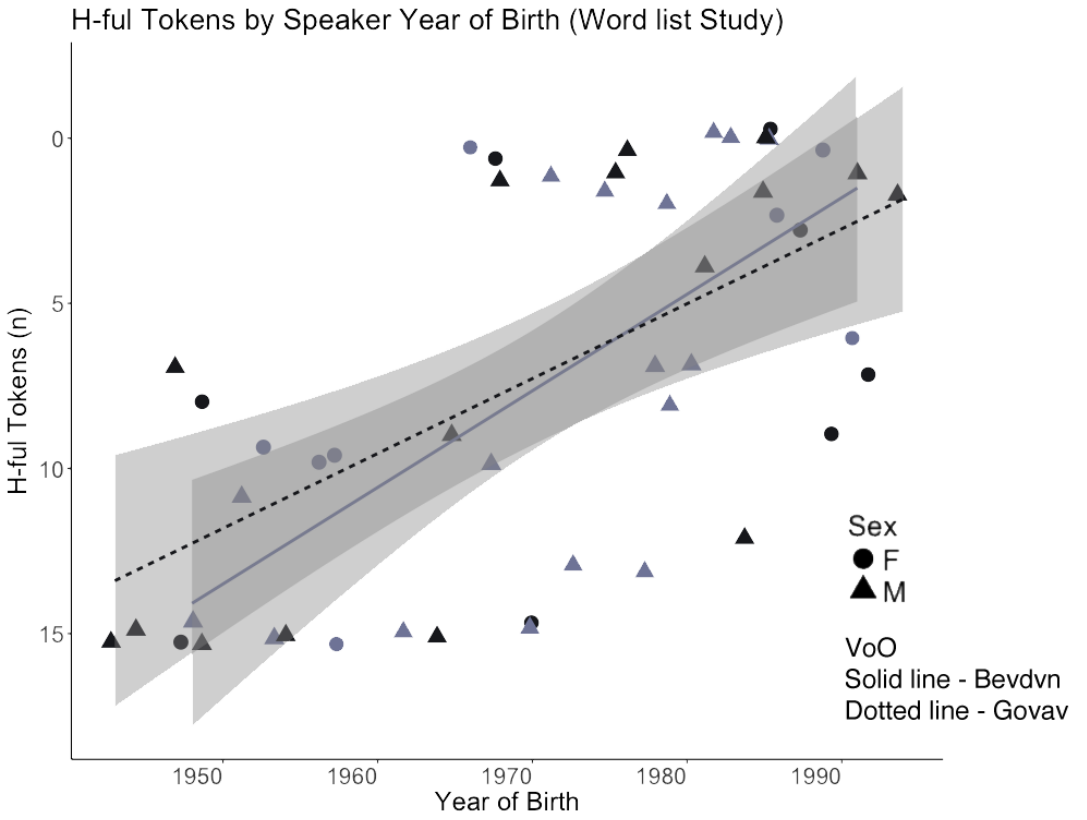

Figure 7.1: Word list study: Number of [h]-ful words, by year of birth of speaker. Each point represents a speaker. Y-axis shows number of [h]-ful realisations, i.e. 15 = categorically [h]-ful. The y-axis has been inverted to be visually similar to the natural speech data results in the following section.

SUM	YoBs	Female		Male		Total
		Speakers (n)	%	Speakers (n)	%	Speakers (n)
15	1943 - 1997	4	8%	9	18%	13
14-10	1951 - 1984	2	4%	5	10%	7
9-5	1947 - 1992	5	10%	4	8%	10
4-1	1968 - 1994	3	6%	9	18%	12
0	1966 - 1989	3	6%	5	1%	8
Total		17	100%	33	100%	50

Table 7.1: Word list study [h]-fulness by gender. Left column indicated number of words produced [h]-fully, i.e. a SUM of 15 is 100% [h]-fulness, while 0 is all [h]-drop. i.e. there are four female speaker who are categorically [h]-ful, which comprises 8% of the word list study participants.

summary of the pilot results is shown in table 7.1.

The results show that [h]-drop rates vary across the board with both [h]-ful and [h]-dropped forms being uttered. A significant portion of the sampled speakers categorically produce [h]-ful forms (13 out of 50 speakers = 26%), whose year of birth spans the entire range of the sample. The majority of these [h]-ful speakers, however, were born prior to 1970. On the left side of the graph are the speakers who categorically produce [h]-less words (8 out of 50 speakers = 16%), most of whom are born after 1975. A large portion of speakers show varying degrees of [h]-drop, but almost half the speakers sampled are producing less than 50% of the words with word initial [h] (i.e. speakers who produce under seven [h]-ful forms out of the 15 total words).

Highly monitored tasks such as reading passages and word lists are said to elicit perceived standard and prestige forms from speakers, but the Nmbo results show this only to some degree. Recall that Nmbo speakers express the "correct" or "true" way of speaking Nmbo is "how old people speak" (section 3), so we might assume the [h]-ful form of words would be understood as the preferred community norm, and that these would be produced at a high frequency in a word list task. Speakers in the word list task do appear to be producing more [h]-ful words, compared to the results from the natural speech data. Some of the speakers in the word list data born after 1980 produce more than 50% of their words [h]-fully, while this is not the case in the natural speech data.[1] So [h]-ful forms appear to be the more "standard" form, but [h]-drop occurs categorically for some speakers. The interpretation of these Nmbo results are that [h]-drop is truly below the level of speaker and speech community consciousness, and that monitoring for standard forms are weaker in this group of speakers as a whole.

A few more patterns can be discerned from this data. Closer inspection at the word level shows that some words are [h]-dropped more than others. This suggest some lexical effect with some words more prone to being analysed as [h]-ful, and other as [h]-less. The words that were least to most [h]-ful are: *hakr* 'boy' (17/50 = 34%), *hkv* 'eye' (18/50 = 36%), *hure* 'straight' (19/50 = 38%), *hamba* 'village' (21/50 = 42%), and *hrare* 'moon' (25/50 = 50%). This suggests that perhaps *hakr* is a particularly [h]-drop prone word in Nmbo. Gender is difficult to interpret here due to the imbalance in sample size, but we can see

[1]The highest retention rate of [h]-ful forms by a speaker born after 1980 is at 24%. Speaker mmggGZ in the NSC shows 8 tokens of [h]-ful varieties out of a total of 33 words.

that proportional to the number of total speakers per gender, there are more male speakers who are categorically [h]-ful compared to female speakers: 27% (9/33) vs 20% (3/15) respectively. In other words, if the [h]-ful variety is the more standard form, we have male speakers producing the standard more than women. This goes against a well-known finding in the English-speaking world that women tend to produce standard varieties in highly monitored tasks (Trudgill 1974, c.f. Romaine 2008:101, Trudgill 1983:162).

Finally, a regression line fit to the data suggests a minor difference in rates of [h]-drop across village of origin. The overall trajectory of [h]-drop by age is the same across speakers of the two villages, but figure 7.1 shows what looks like a village difference in older speakers. Govav speakers showing slightly higher amounts of [h]-drop, until we observe a cross-over where the Bevdvn speakers catch up and over take Govav speakers from 1975. The differences between villages look quite minor, but I show this here because we will see later in the natural speech data a similar tendency. Older Govav speakers show slightly higher levels of [h]-drop than older Bevdvn speakers, while this village-based difference is absent for the younger speakers.

In sum, the word list study suggests that age, word, and gender are all possible candidates for explaining the pattern of variability. It shows that generally older speakers are [h]-ful. Older speakers are identified by the community as speaking the most correct forms of Nmbo, so one may hypothesise that the perceived correct form of word initial /h/ words are the [h]-ful realisations of it. Despite this, some speakers show variability in their production of [h], with many speakers born after 1970 categorically dropping it. There is also evidence of lexical conditioning, with some words more likely to have [h]-drop than others. In order to get linguistically and socially more meaningful insights into this change in progress, we will now turn to utterances of [h]-words in natural speech data.

7.3 Naturalistic Speech Data Result

The data presented from here on were drawn from the Nmbo Sociolinguistic Corpus (NSC). The details of the corpus composition, construction, and annotation, are presented in section 5.3, but the specific details of coding for [h]-drop will be presented now.

The criterion for demarcating the envelope of variation is defined as the occur-

rence of a word initial [h]-word in speech. The variable under investigation is whether the [h] is dropped in that instance or not. All instances of [h]-words identified for this study are lexical nouns. Like the word list study, words are included in the analysis if they are cognate with a Nen word that is /s/ initial, e.g. Nen *sérki* = Nmbo *herge* 'life'. For words that are not cognate with Nen /s/, I verified each word with older speakers of Nmbo to see whether these were pronounced [h]-fully nor not. e.g. Nmbo *hambr* 'leg' is *kaep* in Nen, but was produced [h]-fully by older speakers, and therefore included.

Both linguistic and social conditioning factors that emerged from the word list study were considered and coded. Since the relevant linguistic conditioning factors are not known *a priori*, a few conditions were coded and tested. These conditions can broadly be grouped into the following categories:

- Immediate environment preceding the [h]-word: e.g. preceding phonological environment, including pauses and breaks, the preceding word, the word class of the preceding word;
- Lexical effects pertaining to the [h]-word itself: e.g. lexical conditioning, and frequency.

Following phonological environment, and number of syllables in [h]-word were also explored, but were not found to be significant.

The social variables investigated are speaker year of birth (YoB), Gender, Village of Origin (VoO), and Village of Residence (VoR). Each utterance of an [h]-word was associated with a particular speaker who made the utterance, and these have been collated from across the NSC. The statistical models are logistic regressions, which are a type of generalised linear mixed model used on a binomial variable (e.g. categorical "present" vs "absent" variable).

7.3.1 Linguistic Variables

The NSC yields 1649 tokens across 38 speakers. 942 of these tokens show [h]-drop (57.1% of the data), while 707 do not (42.9%). Figure 7.2 shows the distribution of the ten most frequent [h]-words from the data with a token count higher than 20. The full-list of [h]-words found in the corpus, as well as their frequencies, are given in the appendix (A.16). The rates of [h]-retention and drop per speaker is also provided in the appendix (A.14)

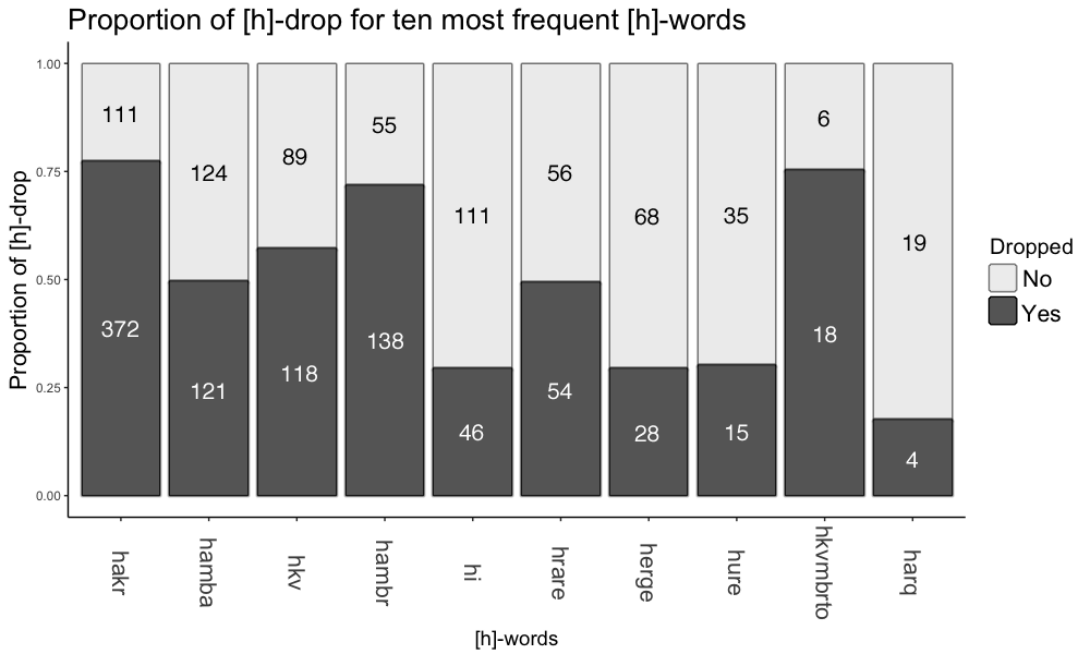

Figure 7.2: Proportional representation of [h] drop for the ten most frequent words. Ordered from left to right, most frequent to less.

Let us consider the linguistic environments of [h]-drop . Firstly, the *preceding phonological environment* was defined by grouping certain phones into natural classes. The distribution of these preceding environments are shown in table 7.2.

Most phonological conditions have [h]-drop occurring at rates higher than 50%. The two conditions with less than 50% of [h]-dropping are nasals (48.7%, n = 95/195), and voiced fricatives (38.2%, n = 13/34). The results present some interesting comparisons to the few variationist studies that look at the linguistic conditioning factors of [h]-drop. As mentioned earlier, preceding liquids and obstruents are reported as favouring [h]-drop in New Zealand English (Bell & Holmes, 1992), which is true for the Nmbo data. Rhotics and voiced stops are both infrequent in the data, but show a high proportion of [h]-drop at 66.7% and 64.7% of occurrences (n = 42/63 and 33/51 respectively). Bell and Holmes's study also report preceding nasals as favouring [h]-drop, and this is also reported for Marshallese English (Buchstaller, 2018). This is not true according to this frequency table 7.2. The data show just under half of the [h]-words following a nasal without the [h] (48.7% dropped, n = 95/195).

The forms of the preceding words were also investigated in order to determine whether phonological conditioning patterns were in fact reflective of common

	Retained (n, %)		Dropped (n, %)		Total (n)
Rhotics	21	33.3%	42	66.7%	63
Voiced Stops	18	35.3%	33	64.7%	51
Open Vowels	80	36.4%	140	63.6%	220
Voiceless Fricatives	11	37.9%	18	62.1%	29
IU Break	223	39.9%	336	60.1%	559
Voiceless Stops	19	41.3%	27	58.7%	46
Closed Vowels	35	41.7%	49	58.3%	84
Mid Vowels	179	48.6%	189	51.4%	368
Nasals	100	51.3%	95	48.7%	195
Voiced Fricatives	21	61.8%	13	38.2%	34
Total	707	42.9%	942	57.1%	1649

Table 7.2: Preceding phonological environment of [h]-words, and the number and percentage of tokens that retain or drop [h] per environment. Nmbo Open Vowels are /æ, ɑ/, Mid Vowels are /e, u/, and Closed Vowels are /i, u/. Semi-vowel [w] is included in the Closed Vowel category.

	Word	Retained (n, %)		Dropped (n, %)		Total (n, %)	
1	yna DEM.PROX	38	41.8%	53	58.2%	91	8.3%
2	tande 1sg.POSS	39	61.9%	24	38.1%	63	5.8%
3	toge ‘child/small’	15	28.8%	37	71.2%	52	4.8%
4	yände 3sg.POSS	26	55.3%	21	44.7%	47	4.3%
5	är ‘man/male’	8	21.1%	30	78.9%	38	3.5%
6	ämb ‘some’	10	30.3%	23	69.7%	33	3.0%
7	bä 3sg.ABS	14	45.2%	17	54.8%	31	2.8%
8	dena ‘like this’	2	11.1%	16	88.9%	18	1.6%
9	mrz ‘girl’	8	50.0%	8	50.0%	16	1.5%
10	ok ‘okay’	7	46.7%	8	53.3%	15	1.4%
10	yao NEG	6	40.0%	9	60.0%	15	1.4%
	Sum of top 10	173		246		419	
	Total (-IU)					1093	

Table 7.3: Top ten most common words that precede [h] words in the NSC. This subset has excluded instances where the [h] word is preceded by an IU boundary.

collocation patterns. Nmbo nouns can be modified by a variety of words (section 4.6), so the number of identical collocations in the data is quite low. For example demonstratives, which precede head nouns, often taking semantic case markers such as the locative *-n*, allative *-t*, the demonstrative ablative *-anma(e)* alter the phonetic form immediately preceding the head (e.g. *yna=no hamba* 'at the village' vs *ynan-mae hamba* 'from the village'). Table 7.3 shows the ten most common words that precede an [h]-word. Of these, two are demonstratives without case marking (*yna* 'this', *dena* 'like that'), and two are possessive pronouns (*tande* 1sg, *yände* 3sg). Three of the top ten words end in [e] while another three end in an open vowel [æ] or [ɑ]. There is no clear indication that preceding word and phonological environment significantly overlap in the data. This reduces the likelihood that collocations with preceding word are affecting [h]-drop rates.

There is, however, evidence to suggest overlaps between *preceding word sub/class* (from here on *word class*[2]), and *preceding phonological environment* overlap.The large number of word classes occurring in the data hinder the ability to make generalisations, so these were grouped into language-internally logical groups. *Modifiers* (MOD) includes numerals, quantifiers, temporal adverbials, the negative particle, proper names, and nouns used attributively. *Grammatical Word* (GRM) includes conjunctions, and TAM particles. *IU* refers to a final IU boundary marker, and includes aborted attempts (e.g. "tha- that boy"). *Discourse Particles* (DISC) refers to non-lexical sounds, such as affirmative "aha" and "e~" , which signals the passage of time. DISC also includes the topic marker *gym* (8). *Demonstratives* (DEM) include spatial and manner demonstratives (4.2.1.3). *Verbs* (V) and *Nouns* (N) are lexical words. Nouns may or may not be case-marked, and this category also includes non-finite verbs. *Pronouns* (PRO) include ergative, absolutive, and other oblique forms of personal pronouns.

The frequency table 7.4 supports the case of significant overlaps between preceding word class and phonological environment. The large number of cells with 0 tokens lies testimony to the fact that the form of the preceding word, and its phonetic forms, are not independent. For example, 55.4% of preceding nasal consonants are due to a preceding verb (n = 108/195), presumably because of

[2]Note that the use of the term "word class" here is not referring exactly to the word classes I described in the sketch grammar, but as a label to group the various words surrounding the [h]-words found in the natural speech data. The label is used for the purposes of this quantitative investigation.

the first person actor suffixes *-n* (singular) and *-m* (non-singular). Just over half of the preceding open vowels are from preceding demonstratives (n = 116/220, 52.7%). Just under half of the mid vowels come from some possessive form (n = 174/368, 47.3%). The majority of preceding rhotics are due to *är* 'man/male', and *mer* 'good' (n = 54/63, 85.7%) modifying an [h]-word.

	IU	DEM	DISC	GRM	MOD	N	POSS	PRO	V	Total
IU	559	0	0	0	0	0	0	0	0	559
Mid Vowels	0	9	12	28	20	81	174	13	31	368
Open Vowels	0	116	0	4	10	29	0	39	22	220
Nasal Consonants	0	9	15	0	2	52	0	9	108	195
Closed Vowels	0	2	15	8	26	16	9	0	8	84
Rhotics	0	0	0	2	54	1	0	0	6	63
Voiced Stops	0	0	0	0	33	1	0	10	7	51
Voiceless Stops	0	4	0	0	0	6	0	2	34	46
Voiced Fricatives	0	0	0	0	1	30	0	2	1	34
Voiceless Fricatives	0	2	11	0	4	8	0	0	4	29
Total (n)	559	142	53	42	150	224	183	75	221	1649

Table 7.4: Frequency table of preceding word (sub)class by preceding phonological environment of [h]-word. Numbers presented are all token counts (n). The large number of 0s between word class and phonological environment shows that many word classes are restricted in their phonological forms, therefore there is a high degree of overlap between the two categories.

If we look at the distribution of word classes alone without reference to the preceding phonological environment, we can see more evidence that the two parameters overlap significantly (table 7.5). For example MOD which shows the second highest rate of [h]-dropping at 64% (n = 96/150) has [h]-drop rates close to that shown by preceding rhotics (66.7%, table 7.2). The rate of 61.3% [h]-drop for DEM is also very similar to the rate for Open Vowels (63.3%, table 7.2). On the other hand, there is some evidence that word class affects rates of [h]-drop independent of phonological environment. A possessive word preceding an [h]-word disfavours [h]-drop, with dropping occurring only 39.9% of the time in the data (n =73/183). Pronouns also show a relatively low rate of [h]-drop (n = 37/75, 49.3%).

The overlap in frequencies between preceding phonological environment and preceding word class shows that preceding environment has an effect. It is not entirely clear which part has more of an effect. Two separate logistic regressions were run to see whether preceding phonological environment or word class is a better predictor of [h]-drop. The models set [h]-drop as a binary depen-

Category	Retained (n, %)		Dropped (n, %)		Total (n, %)
GRM	9	21.4%	33	78.6%	42
MOD	54	36.0%	96	64.0%	150
DEM	55	38.7%	87	61.3%	142
Break	223	39.9%	336	60.1%	559
V	94	42.5%	127	57.5%	221
DISC	23	43.4%	30	56.6%	53
N	101	45.1%	123	54.9%	224
PRO	38	50.7%	37	49.3%	75
POSS	110	60.1%	73	39.9%	183
Total	707	100%	942	100%	1649

Table 7.5: Preceding word subclass of [h]-words in the NSC. GRM = Grammatical Words (conjunctions, TAM particles); MOD = Modifiers (numerals, quantifiers, temporal adverbials, negative particle, proper names, and nouns used attributively); DEM = demonstratives (spatial and manner demonstratives); IU = final IU boundary marker; V = Verbs (inflected); DISC = discourse particles (topic markers, affirmatives; N = Nouns (inflected, nominalised verbs); PRO = Pronouns (personal pronouns excluding the possessive form, indefinite pronouns); POSS = possessives (possessive pronouns)

dent value, with the independent variable for model PE 'preceding phonological environment', and model WC 'preceding word class'. No other variables were included. Model PE does not find any preceding phonological environment as significant, but model WC finds preceding POSS as disfavouring [h]-drop (coefficient = -.91, SE = .35, p< .01, R^2m = .01, R^2c = .71). We can also see this from the numbers in the frequency table 7.5. A plot showing the predicted probabilities of [h]-drop per preceding word class condition shows the probability of the other word classes, and we can see that most word classes predict [h]-drop to occur over 60% of the time (figure 7.3). A log likelihood test comparing the two models found model WC to be a better fit than model PE (chi squared = 0.02), suggesting that the preceding word class is better at explain rates of [h]-drop than phonological environment.

Finally we will briefly consider frequency effects. Figure 7.2 shows the proportion of [h]-drop for the ten most frequent words in the NSC. The left-most word in the x-axis is the most frequent (i.e. *hakr* is the most frequent word in the corpus, *hamba* the second most frequent etc.). While this graph only represents the top ten most frequent words, it provides an impression of the full range of proportional [h]-drop across the entire corpus. The most frequent word *hakr*

Random Effect	Variance	St. Dev.			
Speaker	8.28	2.88			
Word	1.67	1.29			
Fixed Effects	Coefficient	SE	Z Value	p	
Intercept	1.06	0.64	1.66	0.097	
POSS	-0.91	0.35	-2.58	0.010	**
MOD	-0.71	0.39	-1.82	0.069	
DISC	-0.56	0.50	-1.13	0.258	
GRM	0.75	0.67	1.13	0.258	
IU	-0.34	0.30	-1.10	0.271	
N	-0.18	0.34	-0.53	0.596	
PRO	-0.21	0.45	-0.46	0.642	
V	-0.04	0.34	-0.11	0.914	

Table 7.6: Logistic regression model fit summary of [h]-drop YES vs NO, fixed effects of Preceding Word Class, Speaker and Word as random intercepts. IU = final IU boundary marker; V = Verbs (inflected); DISC = discourse particles (topic markers, affirmatives; GRM = Grammatical Words (conjunctions, TAM particles, quotatives); MOD = Modifiers (numerals, quantifiers, temporal adverbials, negative particle, proper names, and nouns used attributively); N = Nouns (inflected, nominalised verbs); PRO = Pronouns (personal pronouns excluding the possessive form); POSS = possessives (possessive pronouns). Observations = 1649, 39 speakers, 20 words. Significance codes: p < .01 = '*' ; R^2m = .01, R^2c = .71

shows the highest rate of dropping at 77% of the data. Incidentally, the tenth most frequent word *harq*, which has a similar phonological form to *hakr* shows the overall lowest rate of dropping with just 17%. It is also interesting to note the similarity of these results with the pilot study. In the pilot study, the most frequently [h]-dropped word is *hakr*, followed by *hkv*, *hure*, *hamba*, and *hrare*. The NSC data also shows *hakr* as the proportionally most [h]-dropped word. This suggests strongly that [h] is dropped more frequently in certain words compared to others.

In order to test the relationship between [h]-drop of lexical item and frequency of occurrence, a Pearson's correlation test was run, investigating ratio of [h]-drop with frequency of occurrence in the data. The lexical items tested for are the words with total token counts above fifty. Frequency of occurrence is set numerically, where a frequency rank of '1' means the most frequent word in the corpus, '2' the second more frequent, and so on. A strong correlation of -0.77 suggests that there is a relationship between lexical item and frequency of use. A visual representation of the the data are shown in figure 7.4.

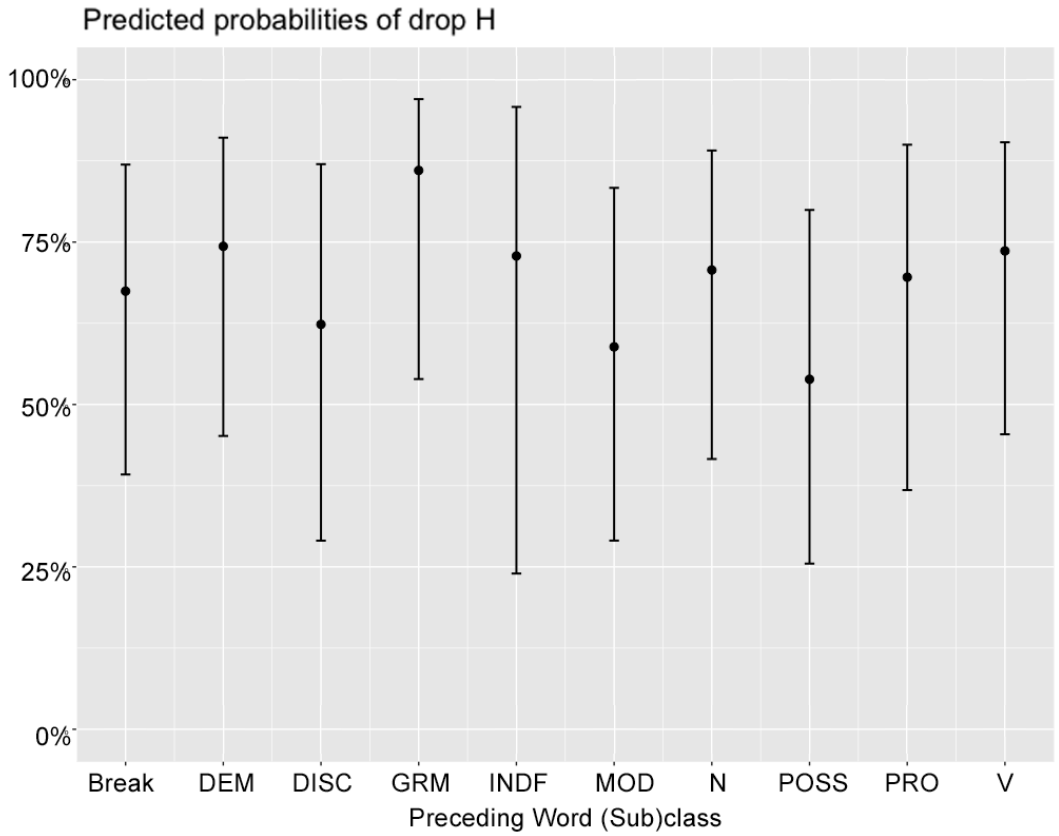

Figure 7.3: Probability of [h]-drop by preceding word (sub)class. Break = final IU boundary marker ; DEM = demonstratives (spatial and manner demonstratives); DISC = discourse particles (topic markers, affirmatives; GRM = Grammatical Words (conjunctions, TAM particles, quotatives); MOD = Modifiers (numerals, quantifiers, temporal adverbials, negative particle, proper names, and nouns used attributively); N = Nouns (inflected, nominalised verbs); POSS = possessives (possessive pronouns ; PRO = Pronouns (personal pronouns excluding the possessive form); V = Verbs (inflected)

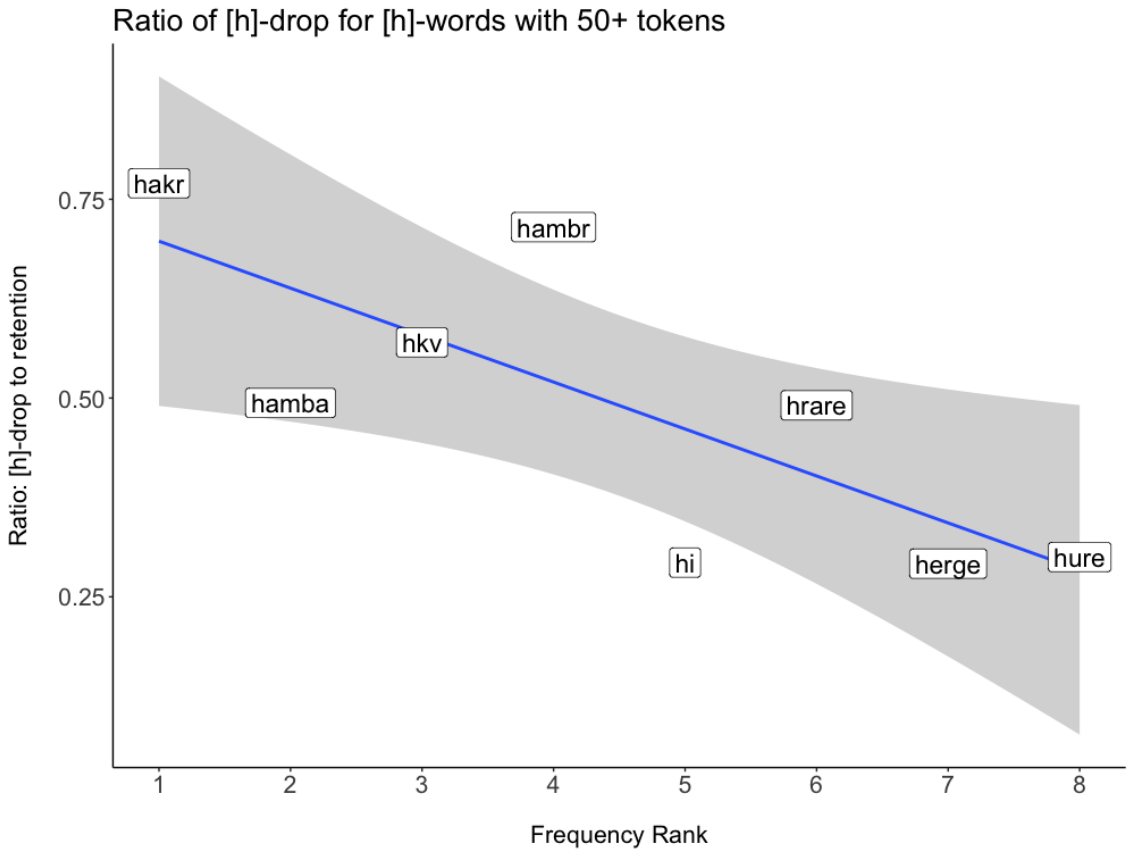

Figure 7.4: Ratio of [h]-drop for [h]-words with 50+ tokens. The lower the frequency rank on x-axis, the more frequent (i.e. '1' is the most frequent word).

While the data shows a strong correlation between frequency of occurrence and [h]-drop rate, it is difficult to assess what this might actually mean. Firstly, while the relative frequency of [h]-words to each other correlates with [h]-drop, these are relatively low occurrence words compared to, say, grammatical words such as demonstratives. High-frequency structures have been suggested as susceptible to phonetic reduction but resistant to analogical change (e.g. Bybee and Thompson 1997), but the Nmbo [h]-words would not qualify as high frequency by most definitions. They are certainly not as frequent as English auxiliary verbs (Bybee and Thompson 1997) or conjunctions (Leech, Rayson, and Wilson 2001:25), *-ing* (Houston 1985, Forrest 2017). Additionally, many quantitative studies of sound change in progress do not show a clear relationship between lexical frequency and phonetic reduction. Neither Dinkin (2008) on Northern American short vowels, or Labov's (2011) extensive work on Northern Cities Vowel shifts find clear frequency effects. Contrary to predictions, Hay, Pierrehumbert, Walker, and LaShell (2015) found that infrequent words are leading in the ongoing changes of BIT, BET, and BAT vowels in New Zealand English, while the high frequency words were lagging. Walker (2012) points out that while Bybee (2000) uses English t/d-deletion as an example of how high frequency words undergo reduction, t/d-deletion is stable variation; not a change in progress phenomenon (p.410). Here we can only say that there is a relationship between token frequency of [h]-nominals and rates of drop, and this is separate from type frequency since there are very few instances of recurring identical collocations in the data.

We have seen in this section that there are many linguistic conditions surrounding [h]-words, but the quantitative results are ambiguous. This is presumably because speaker age is the single best predictor of [h]-drop, and as we will see, this change appears to have progressed so far along as to obscure any linguistic effects that may hint to the origins of this phenomenon. The distributional data suggests preceding word class has an effect on [h]-drop, with possessive pronouns and personal pronouns tending to favour [h]-drop. There is some evidence of preceding phonological environment also having an effect, with preceding liquids and obstruents favouring [h]-drop.

7.3.2 Social Variables

The final conditioning factors to consider are the social characteristics of the speaker. Given that the word list study indicates a strong age effect, it is unsur-

prising to also see a clear age effect in the natural speech data. Here again, age is the single most powerful factor explaining the patterns of [h]-drop.

Figure 7.5 shows that older speakers towards the left side of the graph with lower proportion of [h]-drop. In other words they are more [h]-ful speakers with fewer instances of [h]-dropping compared to the younger speakers towards the right of the graph. A Pearson's correlation test of .82 between year of birth (Yob) and [h]-drop ration suggests a strong relationship, with younger speakers tending to drop [h].

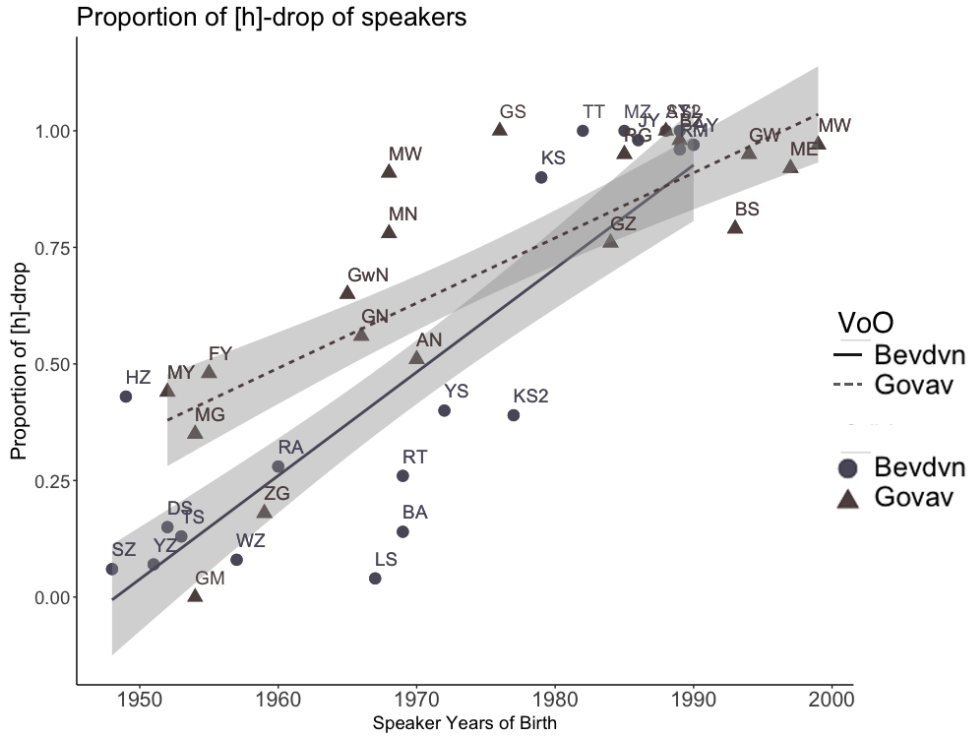

Figure 7.5: Proportion of [h]-drop by speaker, ordered by year of birth. Regression line fit for village of origin: Bevdvn, and Govav.

The other point of notice is that speakers from different villages appear to pattern differently. Three of the four older speakers with higher proportions of [h]-drop (speakers MY, MZ, FY) are all from the village of Govav. Visual inspection of the figure 7.5 with a linear regression fitted to the data, suggests villagers originating from Govav may have higher rates of [h]-drop. The increase in [h]-drop rate also shows a steeper incline for the Bevdvn speakers, while the Govav incline is monotonic. A logistic regression model supports this village difference as significant. A linear mixed model was fit on the speakers over the age of 45 (i.e. born after 1973) with [h]-drop binary 'yes/no' set as the dependent variable, and Gender, Village of Origin as independent variables, and Speaker and Word

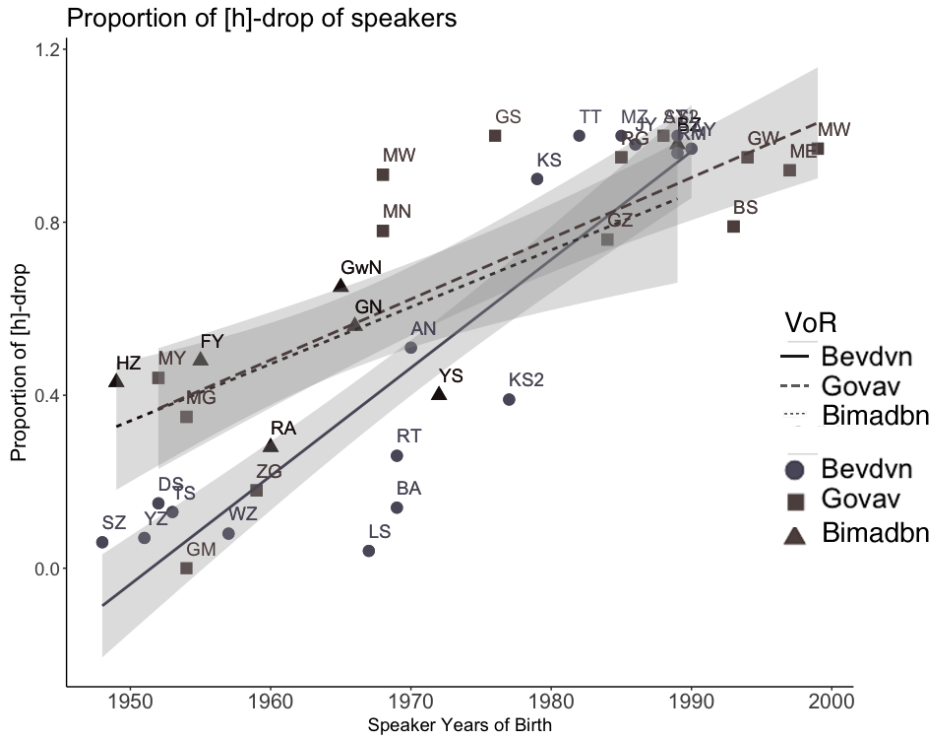

Figure 7.6: Proportion of [h]-drop by speaker, ordered by year of birth. Regression line fit for village of residence: Bevdvn, Govav, and Nen speaking village Bimadbn.

as random effects. Age was excluded as a variable because the effects overpower the model. This model found village to be significant, (SE = .57, p< .03, a = .05, R^2m = .08, R^2c = .42, summary table 7.7).

Random Effects	Variance	SD		
Speaker	1.49	1.22		
Word	1.13	1.06		
Fixed Effects	**Estimate Slope**	**SE**	**Z Value**	**p**
Intercept	-1.48	0.58	-2.54	0.01*
Sex: M	-0.93	0.58	-1.62	0.11
Village: Govav	1.27	0.57	2.21	0.03*

Table 7.7: Results of a logistic regression on speakers over the age of 45, with [h]-drop YES vs NO as dependent variable, Gender and Village of Origin as independent variables, Speaker and Word as random effects. Observations = 1649, 38 speakers, 20 words. Significance codes: p < .01 - .05 = '*'. R^2m = .08, R^2c = .42

Since there is a clear divide between speakers over and under the age of 45, a model was run on the two separate groups to investigate whether there were any differences in conditioning within the groups. We indeed find this to be the case. A logistic regression was run with Preceding Word Class, and VoO as fixed effects,

and Speaker and Word as random effects (summary table 7.8). In the over 45s, we find preceding POSSessives predicted as significant, disfavouring [h]-drop ($p<$ .022, R^2m = .07, R^2c = .42). VoO is also found to be significant ($p<$ .028), with villages from Govav favouring [h]-drop. For the under 45s (speaker n = 18, token n = 672) neither the Word Class nor VoO ($p<$ .1) are significant (full output in appendix table A.30). Preceding phonological environment was modelled for the two groups, but none of the conditions were found as significant in either group. These results suggest very strongly that the linguistic conditioning factors have indeed fallen away for younger speakers, and we have a clear change in progress phenomenon. These results again suggest that preceding word class is more relevant than preceding phonological environment.

Random Effect	Variance	St. Dev.			
Speaker	1.67	1.29			
Word	1.20	1.09			
Fixed Effects	Coefficient	SE	Z Value	p	
Intercept	-1.69	0.60	-2.80	0.005	**
POSS	-0.90	0.40	-2.28	0.022	*
V	-0.08	0.39	-0.20	0.839	
GRM	0.56	0.80	0.70	0.484	
IU	-0.20	0.35	-0.57	0.563	
PRO	-0.19	0.52	-0.36	0.720	
DISC	-0.26	0.57	-0.46	0.644	
MOD	-0.93	0.48	-1.94	0.053	
N	0.04	0.39	0.11	0.913	
Village: Govav	1.32	0.60	2.19	0.028	*

Table 7.8: Logistic regression model fit summary on speakers aged over 45. [h]-drop YES vs NO, fixed effects of Preceding Word Class, and VoO. Speaker and Word as random intercepts. IU = final IU boundary marker; V = Verbs (inflected); DISC = discourse particles (topic markers, affirmatives; GRM = Grammatical Words (conjunctions, TAM particles, quotatives); MOD = Modifiers (numerals, quantifiers, temporal adverbials, negative particle, proper names, and nouns used attributively); N = Nouns (inflected, nominalised verbs); PRO = Pronouns (personal pronouns excluding the possessive form); POSS = possessives (possessive pronouns). Observations = 977, 21 speakers, 20 words. Significance codes: p < .000 - .005 = '**', .01 - .05 = '*'; R^2m = .07, R^2c = .42

The village of residence results are more difficult to interpret since speaker numbers are diluted across three village categories. The token counts were too low to run any statistical modelling on, but upon visual inspection (figure 7.6) we

can see the regression line for Govav residents and Bimadbn residents look very similar to each other. Of the seven speakers who reside in Bimadbn, all but one have married into Bimadbn where they speak Nen. Of those six, two married in from Bevdvn, and four from Govav. We cannot say much more with certainty due to the low token counts, but a suggestive interpretation is that the Bimadbn residing women from Govav are patterning much like the other villagers of Govav.

There does not seem to be any gender effect on [h]-drop. Models were attempted, but most failed to converge, likely due to the low token numbers per speaker for the women as a group. A naive model with Gender as the sole independent variable, and Speaker and Word as random effects, did not find gender to be significant. Visual inspection of a correlation line (figure 7.7) showing that both male and female speakers are affected by age in very similar ways, which supports the lack of Gender as a result in statistical modelling. Female speakers show a slightly higher proportions of [h]-drop by visual inspection, but this is not a significant difference. Recall that the slightly higher proportion of [h]-drop by female speakers was also found in the word list study, but significance tests could not be run on those figures. The similarity between the results of the word list study and this natural speech data study suggests strongly that gender is indeed a non-effect as far as [h]-drop is concerned.

7.4 Discussion and Conclusion

The goal of this chapter was to identify which speakers are at the forefront of [h]-dropping, and to identify any linguistic conditioning affecting [h]-drop. Both the word list study and the natural speech study show there is a strong age effect.

The [h]-dropped form of words are the incoming form, and there is a lot of evidence that change has progressed quite far in the speech community. The word list results are telling, since the genre of highly monitored speech still shows younger speakers producing [h]-less forms of words. The statistical modelling on natural speech data shows age is the single strongest predictor of who will drop [h], with younger speakers likely to drop [h]. The results of the statical modelling which tested for the effects of linguistic conditioning were different for speakers over the age of 45, and under 45. The results show that the older speakers are affected by preceding word class, while younger speakers are not. For the older

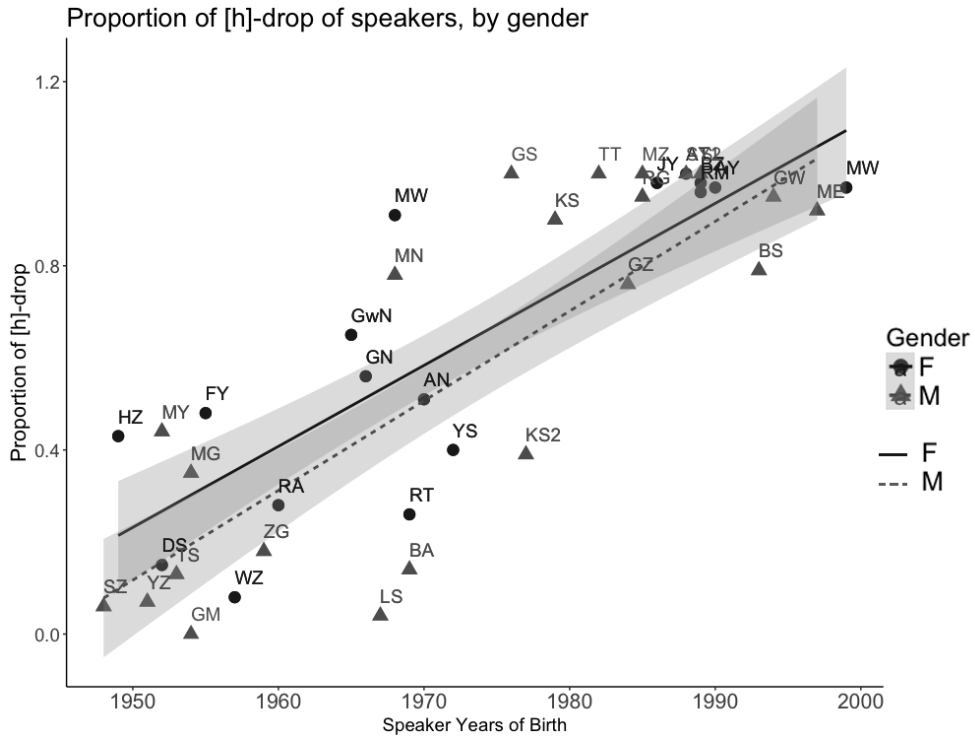

Figure 7.7: Proportion of [h]-drop by speaker, grouped by gender, ordered by year of birth.

speakers, a preceding possessive disfavours [h]-drop, meaning that all the other word classes are more-or-less susceptible to [h]-drop. Phonological environment, when modelled instead of word class, shows no effect on either speakers over or under 45. Since speakers under the age of 45 show no linguistic conditioning, nor any social conditioning affecting [h]-drop, it seems that [h]-lessness has diffused quite successfully through the community.

There is some evidence that the origin of [h]-drop began in the village of Govav. In both the word list study and the natural speech study, the oldest speakers overall show higher amounts of [h]-ful word production. It is, however, the older Govav speakers who have a higher proportion of [h]-drop compared to their peers from Bevdvn. In the naturalistic speech data, the rate of [h]-drop is consistently slightly higher for Govav villagers, until Bevdvn villagers born after the 1980s catch up with similar proportions. The statistical modelling on speakers over the age of 45 show village as significant, while the model for speakers under 45 do not. In other words, the rates of [h]-drop are highly similar for all speakers born after 1972, regardless of their village of origin. The two communities pattern similarly for the younger age bracket, again suggesting a successful diffusion of [h]-lessness through the speech community.

It may be pure chance that Govav village was the origin of the [h]-drop, since there is no reason to assume that a deletion process such as [h]-drop requires some inciting conditions other than the natural processes of articulatory laxity. But the characteristic of Govav as the more multilingual village compared to Bevdvn (section 3.2) makes it the likely origin community for the innovation for two reasons. Firstly, languages spoken in highly multilingual communities have been suggested as having a large feature pool (Mufwene, 2001) from which innovations can arise. Cheshire, Kerswill, Fox, and Torgersen (2011) suggest for multi-ethnic London that high linguistic diversity leads to a heterogeneous feature pool from which new variants are introduced into the speech community. The feature pool, however, seems to require some direct linguistic input into the pool from which speakers then choose features to appropriate. A direct analogy for this Nmbo case would be that some language that does not have word initial [h] for cognate Nmbo words affect the Nmbo forms. There are no specific Nambu branch languages in the areas immediately adjacent to Govav that have altogether lost word initial phones that correspond to Proto-Nambu *s. A weaker version of the feature pool analysis would be that the mere presence of multiple inputs is conducive to phonetic innovations compared to a community which does not have diverse inputs.

The second related factor for why Govav may be the origin of [h]-drop pertains to its location. Govav is ideally placed to spread innovations rapidly should an innovation arise, since it is a multi-link node in the social network of Morehead Area villages. Govav has the highest network connectivity with other villages through marriage. This leads Rueck (2006) to suggest that Govav "is the best place to introduce innovations if one wants them to spread" throughout the network rapidly (p.89). When we look at the links between Govav and Bevdvn specifically, there are five women who have married from Govav to Bevdvn, while there are only two women who has married from Bevdvn into Govav. The greater multiplexity of ties from Govav to Bevdvn might predict a more frequent and/or larger flow of people going from Govav to Bevdvn for short term visits. This is a classic case of diffusion and change, where the younger speakers acquired [h]-dropped lexical nouns as part of their input, either through Govav villagers visiting Bevdvn, or by Bevdvn children following their mothers back to Govav and receiving input there.

Gender does not affect [h]-drop variability, which is of interest given that gender often manifests in variation and change phenomena across various speech commu-

nities ranging from urbanised Sweden (Nordberg & Sundgren, 1999) to regional dialect of K'iche' (Mayan; Guatemala) (Romero, 2009). In the sociophonetic study we have discussed the role of cultural orientation as an explanation for the non-significance of gender (6.4), I take the [h]-drop results here to be something slightly different. Nmbo vowels are high frequency items that are constantly used in speech, and the data suggest that the Nmbo vowel space is a stable system that cannot clearly be characterised as undergoing change. The [h]-drop, on the other hand, is a change in progress below people's consciousness, and is a linguistic item that occurs on a small subset of lexical nouns. The lack of gender variation in [h]-drop suggests that change diffuses through the community through non-gendered connections. The relationship between variation and gender is a mainstay of quantitative language change studies, where gender is understood to reflect complex interactions of various social dimensions in a given speech community; such as social power (Romaine 2008, Labov 2001), social networks and communities of practice (Dubois and Horvath 1999, Eckert 1989, Meyerhoff and Holmes 1999). The implication is, then, that the interaction networks between Nmbo speakers is not so gendered as to manifest differential rates of [h]-drop. Diffusion appears to happen quite rapidly through both genders.

The fact that younger speakers from both Govav and Bevdvn are producing mostly [h]-less varieties of words reveals that this is not a space of linguistic differentiation across the villages. Since lexical differentiation is one of the key ways in which languages differentiate from one another within the Nambu branch languages, one can imagine a scenario where Bevdvn speakers may have retained [h]-ful forms as their village variety in contrast to Govav speakers who were dropping [h]. There are many possible reasons as to why this hypothetical scenario did not take place, but the only thing that can be said with certainty is that the younger Nmbo speakers seem to form a coherent speech community despite the ideological distinction made in community discourse.

In the introduction of this thesis I mentioned that one of the hypotheses pertaining to the high levels of language diversity in New Guinea concern high rates of change (tempo) within speech communities. According to this Nmbo data, the speech community appears to have gone from mostly [h]-ful to [h]-less in the span of the lifetime of the older speakers. The years of birth of the speakers in the natural speech data span forty-five years. While we cannot readily compare studies across time and space to infer rates of change, the Nmbo tempo does not appear particularly remarkable. Studies of consonant change in American communities

show a range of tempos from somewhat similar samples: Velarisation of /l/ in a North Carolinian AAVA speech community is progressing steadily across 69 years of speech data (van Hofwegen, 2011). Voiced/tapped/flapped realisation of medial -*t* in NZE has been steadily increasing over 120 years of ONZE data (Hay & Foulkes, 2016), and usage of innovative variants in the period spanning speaker birth years of 1940-1980 show very steep increases by, say, female speakers (p.311). On the other hand, a study by Stanford, Severance, and Baclawski (2014) shows speakers with near categorical r-less pronounciations in New Englander speech over a 60 year span. The Nmbo data suggests that word initial [h] is well and truly on the way out, and should this happen in the next decade or so, but this tempo of change is reminiscent of better-studied, more populous and dispersed speech communities. In other words, the tempo of consonant loss in Nmbo does not appear particularly faster than other speech communities.

This study has shown that a phonological change in progress appears to have social origins in its inception, and that it diffused from a highly connected village within the village networks of the Morehead area. Gender, like the sociophonetics study, was not significant. The linguistic and social conditionings present in older speakers is no longer visible in younger speakers. The current strongest predictor of [h]-drop propensity is speaker age, with speakers born after 1972 showing a high likelihood of producing [h]-less forms of words. The results of the linguistic conditioning are somewhat ambiguous, but suggests that preceding word class affects differential rates of [h]-drop.

Chapter 8

Study 3: Emergent Topic Marker

> "To understand the origin of [...] syntactic change, then, we must look beyond syntax. During the process of grammaticalization, of course, properly syntactic considerations may predominate more and more, but it is precisely at the blurred margin between the syntactic and the extrasyntactic that the study of syntactic variation is particularly revealing and has the most to contribute." (Sankoff 1988:156)

Study 3 describes and characterises a morpho-syntactically intricate phenomenon of variation in the Nmbo speech community. The linguistic variable is the combination of an adverbial demonstrative *ge* (section 4.3.3.2) followed by a copular verb (DEM+COP) linking an NP to a following clause. The demonstrative *ge*[1] typically functions as a relativiser or general subordinator in complex clause constructions (sections 4.9.6), such as in example 179 below. The verb following *ge* in the relative clause agrees in person/number with the NP, and the TAM of the matrix clause (example 179).

[1]We are mostly concerned with the *ge* form of the demonstrative in this study, but the alternative *gs* is also used in this construction type.

(179) tande **toge-toge-am** [ge t-ng\ne/**tawt**

1sg.POSS **child-RDP-ERG** **DEM** 3pl.A>3sg.U:ipfv.rmpst:and/eat

yna nne...]$_{REL}$ mnd k-ng\aweto/**tawt**

DEM.prox food.ABS vomit.ABS **3pl.A**:ipfv.rmpst:and/masticate

My children that ate this food, vomited.

WSEK1-A20160720-07DroughtWZ 00:01:17.183 - 00:01:21.085

Instances can be found, however, where the copular verb is not showing agreement with its arguments. As we can see in example 180, the copula is frozen into a third-singular and non-prehodiernal (nphd) TAM form, *ym*, and the preceding demonstrative *ge* phonetically leans into this frozen copula. The result is the phonetically and morphologically reduced form *gym* [gʲɐm ~gʲəm ~gɪ.jəm], functioning as an argument topicaliser in a mono-clausal construction.

(180) [An example of a frozen DEM+COP. Hom is talking about when she was a young newly-wed. She explains how she was, at first, scared of her new husband.]

Yndo **gym** tanzo är bérbér-uva

1sg.ERG **DEM+3sg.nphd/be** 1sg.POSSC husband.ABS fear-COM

t\ake/tawn.

1sgA>3sgU:ipfv.rmpst/look

I **[TOPIC]** looked upon my husband with fear.

WSEK1-B20150813-02HomHZ, 00:02:50.880 - 00:02:53.755

The development of demonstratives into topic constructions is a well-known areal tendency of New Guinean languages (De Vries 2006, 1995, Reesink 1994). Known as *thematising markers* in the Papuanist literature, constructions of this kind have been found across the New Guinea mainland from Eipomek in the west (Trans-New Guinea, Mek; Indonesia. Heeschen 1998), Korafe in Milne Bay to the east (Trans-New Guinea, Greater Binanderean; Indonesia. Reesink 1994). Even Paluai, an Oceanic language of Manus Island, has demonstratives as topicalisers and markers of definiteness (Austronesian, Admiralties; PNG. Schokkin 2014).

Thematising functions include recapitulation, resumption of preceding themes in discourse, linking distributed arguments across clauses (Heeschen 1994:57-65), or chaining paragraphs (Farr 1999:337-342). The pattern in Nmbo is another example of this well-known New Guinean phenomenon.

The goal of this chapter is to investigate the linguistic and social patterns surrounding the loss of DEM+COP morpho-syntactic agreement. The question is whether there is evidence of an emergent topic marker from a subordinating constituent. This Nmbo phenomenon fits with observations that mono-clausal topic/focus constructions often arise from bi-clausal constructions (e.g. Harris and Campbell 1995:152-156), which is also related to the semantics and syntax closeness of relative clauses and topic/focus constructions (e.g. Schachter 1973). Indeed, the emergence of topic markers from relative clause constructions has been noted as a common path of grammaticalisation by Heine and Kuteva (2002b:96). Haiman (1978) also famously made the case that conditional clauses and topic constructions are the same in kind due to semantic and formal similarities, using examples from Hua (Trans-New Guinea, Kainantu-Goroka; PNG) to make his case.

The two phonetic-phonological studies in this thesis thus far were investigated using variationist methodologies, but extending this method to the investigation of DEM+COP morphosyntactic variability is significantly more challenging for a number of reasons. I will shift gears slightly and borrow elements from grammaticalisation approaches to describe the conditions of DEM+COP agreement loss as a case of morphosyntactic *decategorialisation*. I will begin this chapter by describing DEM+COP and its variable context (section 8.1). I will then present the rationale (section 8.2), and outline the contextual model of grammaticalisation used in this study to investigate DEM+COP decategorialisation (section 8.3). Evidence of DEM+COP decategorialisation as incipient grammaticalisation of a topic marker will be presented (section 8.5), and then a short comment will be made on the lack of detectable social aspects (section 8.6).

The relevant sections of the sketch grammar for this study are the adverbial demonstrative section (4.3.3.2), case marking (4.4.4), noun phrase structure (4.6), verbal morphology (4.5), inflectional TAM (4.7), simple clausal syntax and argument marking (4.8), and relative clause structure (4.9.7).

8.1 The DEM+COP Constituent and Construction

The linguistic item under investigation is a DEM+COP constituent following a lexical or pronominal NP, which is further followed by a clause that minimally contains an inflecting V. When agreement is functioning at all levels, the copula of the DEM+COP cross-references the NP and functions as its main verb, while also agreeing in TAM with the following V. The construction can be abstracted as in figure 8.1:

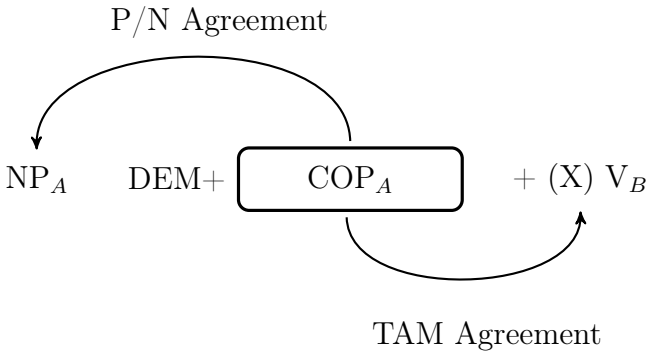

Figure 8.1: Agreement pattern of DEM+COP. The copula component agrees with the NP person/number (P/N) and V TAM. X = NP or other preverbal constituent.

Example 181 shows the position of DEM+COP in a fully agreeing construction. To the left of DEM+COP lies the NP, which agrees in person number with the copula (in this case 1sg). The final constituent of the utterance is the main verb, inflected for imperfective primordial (ipfv.prim) TAM. The copula agrees with the main verb, as it is also inflected for primordial TAM:

(181) [Maiwa is being interviewed about the time he was a young man.]

a.	ynd	**ge w/maro**,	dev-am	ge	w\asen/dam.
	1ABS	**DEM 1sg.prim/be**	father-ERG	DEM	3sg.A>1sg.U:ipfv.prim/look.after

Me **[TOPIC]**, father used to look after me.

WSEK1-G20170707-03MY09Finalhm, 00:00:26.492 - 00:00:28.914

A caveat must be added to the statement that the DEM+COP agrees with V TAM. As noted in the verbal morphology section of the sketch grammar (sec-

tion 4.5.8), the copula can only express a subset of TAM specifications that are available to ambifixing verbs. If the main verb is not in imperfective TAM that the copula can fully agree with (i.e. non-prehodiernal, anterior past, primordial past, remote past), the copula tries to "best match up" semantically with the V TAM. In example 182a the DEM+COP is in the nphd TAM while the main verb is in the perfective past (pfv.pst). As the copula does not form a perfective paradigm, the best semantic match is the nphd tense. Example 183 shows something similar with the best semantic match to the main verb perfective preterite (pfv.pret) TAM being the remote past (rmpst) TAM. Table 8.1 shows the best-match correspondences between DEM+COP and main verb TAM.

(182) [Taqm is telling a mythological story.]

a. niñ-am **ge y/m**, mrz di **y\ane/sa**...
demon-ERG **DEM 3sg.nphd/be** girl.ABS ALR **3sg.A>3sg.U:pfv.past/grab**
The demon **[TOPIC]**, it grabbed the girl...

WSEK1-B20150805-01NinyiTS, 00:05:34.120 - 00:05:36.500

(183) [Alqi is talking about how her pregnant mother gave birth to her at a garden place.]

a. Ktanmae brakean ama **ge t/maro** ktanma
from.there pregnant mother.ABS **DEM 3sg.rmpst/be** from.there
n-n\owano/i...
3sg.A:pfv.pret:ven/arrive
From there the pregnant mother **[TOPIC]**, she arrived there [to the garden place]

WSEK1-G20170617-01Alqi02MQ, 00:01:24.871 - 00:01:28.576

The DEM+COP can be analysed as part of a general subordinate clause construction. The subordinate clause verb in Nmbo can be any inflecting verb including the copula, which is preceded by the demonstrative *ge* or some other subordinator (section 4.9.7). The subordinate copula agrees in person/number

Verb TAM	Copula TAM	Example Verb *wavroh* 'to do X'	Corresponding Copula Form
Imperative (ipfv.imp)	nphd	tavrota	ym
Non-Pre Hodiernal (ipfv.nphd)		yavrot	ym
Past (pfv.past)		yavro	ym
Future (pfv.fut)		yavronga	ym
Yesterday Past (ipfv.ypst)	ypst	tavrot	tm
Proximate Inceptive (pfv.proxin)		tavro	tm
Primordial (ipfv.prim)	prim	yavrotam	ymaro(n)
Preterite (pfv.pret)	prim or rmpst	yavroy	ymaro(n)/tmaro(n)
Remote Past (ipfv.rmpst)	rmpst	tavrotaw	tmaro(n)

Table 8.1: Best-match for TAM specifications across copula and ambifixing verbs. The 'Example Verb' column shows a fully inflected verb *wavroh* 'to do' with 3sg actor and undergoer (except for the imperative form). The 'Corresponding Copula Form' column on the far left shows the 3sg form of the copula which best-matches in terms of TAM.

with the sole argument, and also agrees in TAM with the matrix clause verb. The argument of the subordinate clause will also be cross-referenced on the main verb either as S, A, or U. The expression of the whole clause is something akin to "the NP that is there V_R, does V_M." (example 184). When the subordinate clause verb is a copula, the meaning is something like "the NP that is there$_R$, did V_M" (example 185).

(184) [Zoga is talking about the time an Australian man came to the area to recruit workers for a diving job. The sole argument of the subordinate clause is also A of the matrix clause verb.]

a. [yna site är markai **ge** **t-n/maro**]$_R$
this white man.ABS white.fella DEM 3sg.nphd/come
[n-n\ovar/yng Sigabadru hamba=t.]$_M$
3sg.A:pfv.pret:ven/arrive Sigabadru village=ALL
[This white man that came] he arrived at Sigabardu village.

WSEK1-G20150826-03AusWorkZG, 00:04:06.430- 00:04:15.545

(185) [Maiwa is talking about how he was married as a young man.]

a. [ynd **ge** **w/maron**]$_R$ [dev-am ge w\asend/am]$_M$

1ABS **DEM** **1sg.prim/be** father-ERG DEM 3sg.A>1sg.U:ipfv.prim/marry

[It was I] that father married off.

WSEK1-G20170707-03MY09Finalhm, 00:02:06.553- 00:02:08.033

The DEM+COP construction in Nmbo has a unique intonation contour that shows its bi-clausal form. The two parts A and B form two distinct phonological units. Part A which contains the DEM+COP unit ends with a falling intonation contour, with the pitch dropping over the DEM+COP unit in particular. Part B then begins with with a pitch reset before ending with a final fall (figure 8.2). It is also common to have a pause intervening between the two clauses, in which case the pitch drop over the DEM+COP is just as, if not more, acute (figure 8.3).

In the usage of some Nmbo speakers, the copula of the DEM+COP loses agreement with its associated constituents. We can only see this disagreement when the NP is not 3sg, and/or the V is not nphd. In example 186 we see TAM agreement, but loss of person/number agreement where the DEM+COP takes the form *gym* despite the preceding dual number NP. In example 187 the NP and COP are in agreement, but with V TAM disagreement. The final example 188 shows both NP and V TAM disagreement, where the DEM+COP *gym* can only be interpreted as a topic marker.

(186) [Person/number disagreement.]

a. yna sombwi är **gym** dena ḡav=ut

DEM.prox two person **TOPIC** DEM.mn face=ALL

e\wake/wn

1sg.A>3du.U:ipfv.nphd/see

The two people **[TOPIC]**, I see them like this.

WSEK1-G20170630-01GW03Casshm, 00:02:67.064- 00:02:67.705

(187) [TAM disagreement.]

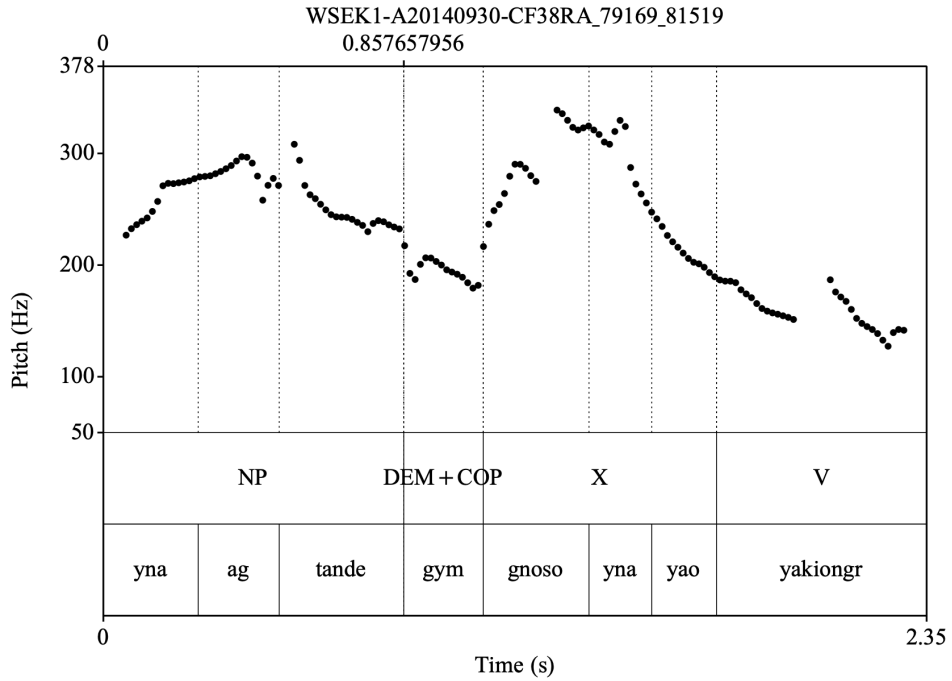

Figure 8.2: Pitch contour of DEM+COP construction, with pitch dropping over the DEM+COP constituent before resetting at the onset of clause B. The utterance can be translated as "This coconut of mine here [TOPIC] now it does not stand." (Ruscien, WSEK1-A20140930-CF38RA)

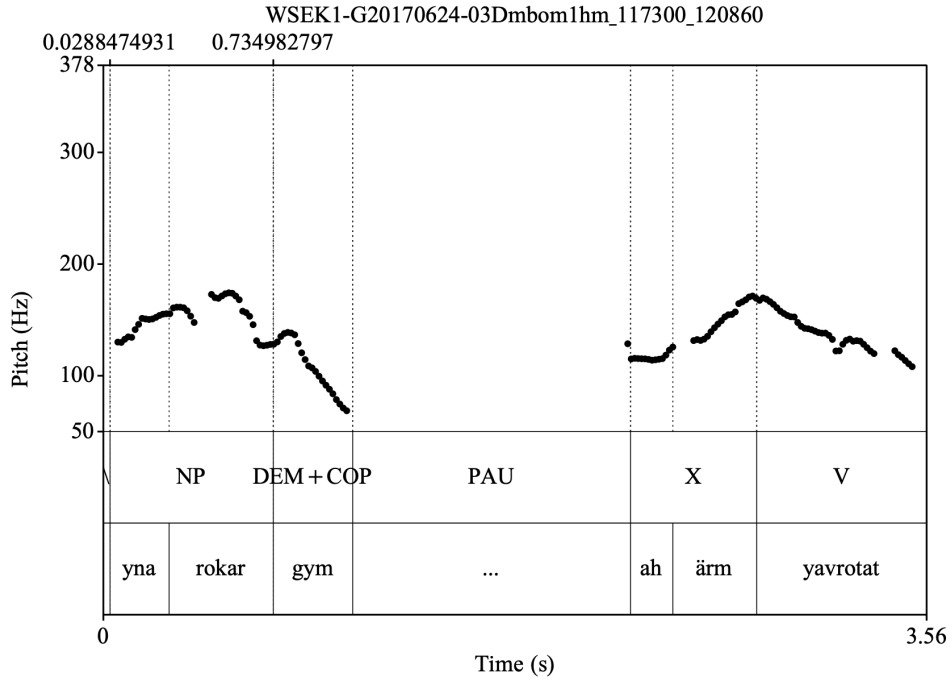

Figure 8.3: Pitch contour of DEM+COP construction, with pause between part A and B. The utterance can be translated as "This object here [TOPIC] ah, people use it." (GimaM, WSEK1-G20170624-03Dmbom1hm)

a. Waria **gym**, dena uñhar-är tmaro,
Waria **TOPIC** DEM.mn to.lie-AG.NMZR 3sg.rmpst/be
Waria **[TOPIC]**, was a liar.

WSEK1-B20170726-01Yavs03Wariahm, 00:01:38.330- 00:01:40.170

(188) [Both person/number and TAM disagreement.]

a. bmo **gym**, evh got-an nne bä ynehai.
2sg.ERG **TOPIC** time core=LOC food 3ABS 2sg.A>3sg.U:pfv.pret/cook
You **[TOPIC]**, were cooking the food at noon.

WSEK1-A20170724-Rusien01Ypsthm, 00:01:04.592- 00:01:06.592

I have opted for the term *topic* in describing the function of decategorialised DEM+COP. As mentioned, Papuanist terminology in the 1970s-1990s uses *theme* for similar constructions (in contrast to *rheme*, a concept from the Prague School imported into Systemic Functional Grammar, c.f. Cummings 2005:129, Hasan and Fries 1995). Nonetheless I have chosen *topic* since it appears to be the most general out of a number of other related terms such as focus and prominence. Topic is defined by some as concerning the discourse "aboutness condition" (eg. Kuno 1973, Chafe 1976 "discourse topic"), or as a mechanism of framing referents as important for the comprehension of discourse; what Reesink (1994) calls "domain-creating constructions". According to Gundel and Fretheim (2008), the term *topic* is reserved for "what the sentence is about" while *focus* refers to "what is predicated about the topic" (p.1). The term focus is, however, also used in a general sense such as by Dik (1989) who suggests focus is the information which the speaker considers most pertinent for a listener to incorporate into the discourse context (p.277). The disjunct between the various functions ascribed to terms such as topic and focus in different case studies is, as Haspelmath (2010) point out, due to the gap between language-particular descriptive formal categories and cross-linguistic comparative concepts. Either way, the precise definition of topic is not particularly pertinent to this stage of investigating Nmbo DEM+COP agreement loss. We will see below that DEM+COP and *gym* are used in a variety of discourse functions that can generally be characterised as the speaker drawing attention to an important argument.

The DEM+COP and *gym* are used for the following three functions:

- The DEM+COP topic marker presents a new entity NP into the discourse;
- The NP preceding DEM+COP is re-emphasised;
- The NP preceding DEM+COP is contrasted against some other clausal topic referents in the preceding discourse

DEM+COP can highlight new and/or given information in the discourse. Discourse given-ness pertains to "a linguistic expression and a corresponding non-linguistic entity in the speaker/hearer's mind" (Gundel and Fretheim 2008:2, also Chafe 1976:30-33) that is assumed or established as the clause topic at any moment in discourse. As discourse progresses, topics often shift. Topic markers function to identify which discourse referent is already established (given), and may also signify the presence of a new discourse referent (newness). The Nmbo DEM+COP construction often has such presentative functions, with the DEM+COP constituent following an NP whose referent is new in context of the discourse (example 189).

(189) [Smako is retelling a story about how he discovered his passion for music while he was in school. In the following segment he is relaying the moment a female classmate asked his aunt about his skills. The TOPIC introduces the referent of "this kid" to the interlocutor.]

a. foyo, yna mrz k-ng\owaito/tao, mwidädem
after.that DEM.prox girl.ABS 3sg.A:ipfv.rmpst:and/follow aunt.ABS
t\mdo/nga,
3sg.A>3sg.U:pfv.b/tell
After that [short performance I gave], this girl followed [me], she told aunty,

b. "Yna toge **gym**, nmbo wah-dvnär y/m?"
this child.ABS **TOPIC** what sing-AG.NMZR 3sg:nphd/be
"This kid [**TOPIC**], what, he is a singer?"

c. “Yao, bä dena bädi n-ng\awatambne/t.”

NEG 3ABS DEM.mn INTS 3sg.A:ipfv.nphd:and/learn

“No, he is learning.”

B20170627-03SmakoSL02Finalhm, 00:01:32.005 - 00:01:37.720

DEM+COP constructions often re-emphasise the topic NP, which corresponds somewhat to the notion of given-ness mentioned above. In Nmbo, the referent of the topicalised NP can be mentioned in discourse immediately preceding the construction, and the DEM+COP serves to emphasise that topic again:

(190) [Jacklyn is narrating about the action taking place in a picture from the Cassowary Picture Task. The NP *orodvnär* 'thief' occurs in the clause preceding the TOPIC clause.]

a.	orodvnär	äuya	n\ok/awt.	orodvnär-m
	thief.ABS	cassowary.ABS	3sg.A:ipfv.nphd/grab	theif-ERG
	gym,	äuya	y\ane/.	
	TOPIC	cassowary.ABS	3sg.A>3sg.U:pfv.pst/get	

The thief is grabbing the cassowary. The thief **[TOPIC]**, he got the cassowary.

A20170611-02RTJY04Retell, 00:01:06.538 -00:01:11.475

(191) [Lazarus is recounting the time the governor of the South Fly district visited Arufi village a few days prior to this recording. He is talking about how the villagers of Arufi were trying to explain to the governor that the money set aside for the school children's sports program had been stolen.]

a.	toge-toge	gs	y-ng\awambn/	kembone,
	child-REDP	DEM	3sg.A>3pl+.U:pfv.pst:and/play	game

The children played many games,

b.	ämb	sonzärbwe	ktano	de	y\nfo/yng...
	some	money	from.there	ALR	3sg.A>3sg.U:pfv.pret/give?

He [the governor] gave them some money for that.

c.	yna	kembone	mani	**gym**...	sonzärbwe	**gym**
	DEM.prox	game	money	**TOPIC**...	money	**TOPIC**
	de	y-ng\ui/tamnd		o,		
	ALR	3pl.A>3sg.U:ipfv.prim:and/steal		CONJ		

This game money [**TOPIC**]... the money [**TOPIC**] they [unidentified thieves] stole it,

d. yve emo y-ng\ane/i oro-h-ye.

3nsg.OBL who.ERG 3sg.A>3sg.U:pfv.pret:and/get steal-INF-DAT

they [the villagers of Arufi] don't know who stole it.

B20150928-MPVisitLS, 00:03:08.826- 00:03:19.153

Finally, DEM+COP often marks the topicalised constituent to contrast against previously established discourse topics.

(192) [Fangore is telling a story about two animals: the butterfly and the tortoise. The rows a and b concern the butterfly as the actor. In clause c, the TOPIC shifts attention to the tortoise.]

a. Dardar t-n/maro yänzu nänzi yta yu

butterfly 3sg.rmpst/came 3sg.POSSC banana FRUS place

mna... mna yta y\uvu/i...

DEM.fw... DEM.fw FRUS 3sg.A>3sg.U:pfv.pret/taste

The butterfly came to this place and tried his banana... he tried tasting it.

b. yao, mämä mnde=t t/maron!

NEG bad taste=ALL 3sg.rmpst/be

No, it was bad tasting!

c. malind-am **gym**... yände nänzi wao t\ne/tao,

tortoise-ERG **TOPIC**... 3sg.POSSC banana ripe 3sg.A>3sg.U:ipfv.rmpst/eat

mé mnde=ro .

CONT sweet=RST

As for the tortoise... he ate his ripe banana, it was very sweet.

G20170724-01FY10Farthm, 00:02:28.610 - 00:02:43.592

8.2 Change from Variability: Variationism and Grammaticalisation

Unlike the two preceding quantitative studies presented in this thesis, this chapter will take a grammaticalisation approach to studying the decategorialisation of DEM+COP. In this section I will present the rationale for this change in approach, and then explain Heine's contextual model of grammaticalisation (2002) to characterise the current state of variation Nmbo DEM+COP appears to be in.

Grammaticalisation has been defined in many ways since the term was first used by Meillet (1912), but a common definition is that grammaticalisation is a long-term process of structural change where a single linguistic item changes from a less grammatical to a more grammatical status (Kuryłowicz 1965:52, Heine and Kuteva 2002b:2). It has been pointed out that there is no theoretical or explanatory power behind the processes identified in grammaticalisation (e.g. by Campbell 2001, Newmeyer 1998, Fischer 2011), and there are no obvious causal explanations for how and why these kinds of changes occur (c.f. Heine 2003). There is increasing recognition, however, that the processes involved in phenomena identified as grammaticalisation are broader in scope. Processes involved in grammaticalisation such as analogy and semantic/pragmatic extension are argued as being involved in other kinds of linguistic change phenomena (Frajzyngier 2008, Heine 2003:475, Smith, Trousdale, and Waltereit 2015:1), as well as being processes that are part of a general cognitive capacity (i.e. extra-linguistic, c.f. Breban, Vanderbiesen, Davidse, Brems, and Mortelmans 2012:5-9). Hence Diewald and Smirnova (2010) describe grammaticalisation as "composite in nature" with a number of other processes (p.98), while Bybee (2010) describes grammaticalisation as an epiphenomenon of domain-general cognitive processes. Despite the still subtle way the term is used, *grammaticalisation* is a useful shorthand for encapsulating a common tendency for change in linguistic structures to undergo compression in form, accompanied with extension of meaning into once separate domains.

The following three points are often accepted as common features of grammaticalisation phenomena: reduction in the realm of phonetics, loss of morphosyntactic properties (decategorialisation), and semantic bleaching (Heine 2003:579, Heine and Kuteva 2002a:378). Phonetic reduction of a linguistic feature results in the

overall compression of the acoustic form of a word (e.g. *be going to* > *gonna*, [gʌnə]). Accompanying phonetic reduction is the loss of morpho-syntactic properties that are characteristic of the linguistic item. Morpho-syntactic compression includes characteristics such as a shift of a linguistic category from a "major category" or word class to a "minor" one (e.g. Hopper and Traugott 1993:104), as well as the loss or neutralisation of morphological markers and syntactic privileges (Hopper 1991:22). An example of decategorialisation is the applicative prefixes in some Timor Alor Pantor languages, which are thought to be serial verbs that have lost their inflectional properties and now behave like a prefix or preposition (Klamer 2018). Desemanticisation is the gradual loss of older semantic associations, as can be seen with the Tok Pisin *laik* (derived from English *like*) now used with purely temporal proximative uses, as opposed to desiderative uses (Romaine 1999).

Variationist and grammaticalisation approaches both start with the recognition of the inherent variability of language, and that future linguistic forms may emerge from synchronic micro-variations. Both approaches are fundamentally evolutionary in model (Croft 2000:16), and share many assumptions about how linguistic structures change and develop over time. Both fields emphasise change as emerging from language use (e.g. Brinton and Traugott 2005:33-34, and Bybee et al. 1994 for grammaticalisation), with variationists particularly emphasising the unmonitored vernacular as the locus of change (e.g. Poplack and Torres Cacoullos 2013:9, Tagliamonte 2012:2). Both fields study the variable use of a single linguistic form (e.g. what Traugott (1995) calls *layering* in grammaticalisation, and Sankoff (1988) calls *form-function asymmetry* in variationism). Both assume the gradual nature of change, such as Lichtenberk (1991) and the *principle of gradual change in function* for grammaticalisation, and what Labov (1994) refers to as *gradualism* of change (pp.23-24) .

Despite the conceptual similarities, variationist and grammaticalisation studies have a fundamental difference in methodology. Variationist methodology has from the very outset emphasised the identification of the total pool of utterances in which a feature varies (Tagliamonte 2012:10), known in Labovian parlance as *the principle of accountable reporting* (Labov 1972c:72). This requires the identification of contexts where the variable in question *may* arise, in addition to the contexts where it does arise. As Poplack (2011) points out, many grammaticalisation studies are form-based, meaning less attention has been paid to the circumscribing of the variable context (p.212). As a way of an example,

a variationist would look at how modal expressions of future can be expressed variably by *will* vs *gonna*, while a grammaticalisation study will look at how *gonna* as a future marker has emerged from the context of *going to* as a verb of motion. Circumscribing possible occurrence sites of variation as well as actual occurrence sites is a particular challenge for non-phonological variables (Lavandera 1978), and this has likely contributed to the relative lack of morpho-syntactic and semantic variables in variationist studies (Nevalainen and Palander-Collin 2011:119). The marrying of variationist methodology and grammaticalisation approaches is also a challenge due to the sheer amount of data needed to satisfy the methodological rigour of both, but such studies are gradually becoming more common (e.g. Petré and van de Velde 2019, Nevalainen and Raumolin-Brunberg 2017, De Smet 2016, Torres Cacoullos and Walker 2009).

The strength of grammaticalisation studies has been in the central consideration of semantics and pragmatics as a mechanism of change. While variationists have considered macro-scale issues such as understanding the spread of linguistic innovations through populations (c.f. Nevalainen, Raumolin-Brunberg, and Mannila 2011, Nevalainen and Palander-Collin 2011), grammaticalisation studies have tended to engage more with the linguistic issues of structural and semantic ambiguity. That is, grammaticalisation research has been more comfortable in considering and theorising how linguistic forms are construed to new contexts via semantic analogy and/or pragmatic inference (e.g. in the works of Traugott 1989, 2010, Traugott and Dasher 2002, Diewald and Ferraresi 2008). There is, however, no reason *a priori* why the tendencies of the two fields should be this way. Especially so since early discussions in variationist research noted how discourse neutralises distinctions in referential value or grammatical function of linguistic variables (Sankoff 1988:153).

Both variationism and grammaticalisation are concerned with the linguistic and pragmatic contexts of variable linguistic phenomena, but grammaticalisation more explicitly emphasises the highly linguistically ambiguous contexts as the loci of innovation. To use the words of Klamer (2004), "Since category change does not take place overnight nor in clearly distinct stages... we expect to find cases of category change 'on the go'. Cases to look for are those where in one synchronic stage a single item has various functions, interpretations and combinatorial properties, and occurs in distinct but co-existing structures, as well as a sizeable number of ambiguous contexts." (p.300). In the following section I will outline the model and terminology of a contextual model of grammaticalisation

(Heine, 2002) as a way of interrogating the various contexts of DEM+COP decategorialisation, including those ambiguous contexts where variation is thought to arise. The model is a useful frame to understand what appear to be multiple intermediary stages of decategorialisation in the current Nmbo data, and sets up the use of multiple conceptual tools to investigate different parameters of DEM+COP compression.

8.3 A Contextual Model of Grammaticalisation

Heine's *contextual model of grammaticalisation* (2002) provides a framework to trace the concurrent progress of linguistic compression across multiple domains. The verbal description of the various stages necessarily requires one domain of compression to precede the others (e.g. describing phonetic reduction prior to morpho-syntactic decategorialisation), but no claims are made as to the actual order of the process.

The various stages of the model illustrate how linguistic contexts provide the opportunities for phonetic/semantic/morphosyntactic ambiguities to arise. At any given point in time, a linguistic construct will have a source meaning A. Through frequent contextual use source meaning A will be extended by speakers into new contexts. This extension occurs through speaker's inferences of meaning A, through metaphor or metonymy, abstraction or concretization. Over time the linguistic construct A will have another meaning B, which is derived from the original source meaning A.

Stages	Context	Resulting Meaning
I Initial Stage	Unconstrained	Source meaning
II Bridging Context	There is a specific context giving rise to an inference in favour of a new meaning	Target meaning foregrounded
III Switching Context	There is a new context which is incompatible with source meaning	Source meaning backgrounded
(IV Conventionalisation)	The target meaning no longer needs to be supported by the context that gave rise to it; it may be used in new contexts	Target meaning only

Table 8.2: Reproduction of Heine 2002:86, with the addition of parentheses for optional stage IV.

The first stage is the *initial stage* where the linguistic construct is near cate-

gorically interpreted as having meaning *A*. Then comes the intermediary stage of *bridging context* (Evans and Wilkins 2000), also known as a *critical context* (Diewald 2002). In this stage the target meaning *B* is an ongoing plausible interpretation of source meaning. This is followed by the *switch context*, where the target meaning is separated from older meanings, and the source and target have incompatible or conflicting properties (Heine 2002:86). Finally comes *conventionalisation*, where the switching context does not support the context of the target meaning. The target meaning can stand alone. Conventionalisation as a stage is not pertinent to a synchronic study of variation, so this study will focus on the initial stage, the bridging, and the switching contexts.

A concrete application of this model to the grammaticalisation of Solomons Pijin transitive marker *-im* will look something like the following. The initial stage involves input from the lexifier language English, where *him* is used as the human third singular object pronoun, e.g. "I didn't see him". In the bridging context, as the form is reduced down to *im*, this item could be analysed as phonetically leaning on the preceding verb. The semantic context allows for the interpretation of the *(-)im* as a human third singular object pronoun (source meaning), but an equally possible interpretation is that of a general object marker (target meaning). In such an interpretation, *im* would be analysed as an agreement marker of the dropped/elided/ellipsed third singular referent. The example below shows two possible interpretations of the same clause:

(193) [Bridging Context: Two possible analyses of the one clause.]

a. Mi no luk **im**.
1sg.ERG NEG see 3sg.ABS
SUBJ OBJ
I didn't see him/it.

b. Mi no luk**-im**.
1sg.ERG NEG see-3sg.OBJ
SUBJ
I didn't see him/it.

(Keesing 1991:319)

In the switching context there are examples of *(-)im* being used in contexts that cannot be interpreted as a third singular (Keesing 1991:318-319, also Sankoff 1993:120 for Tok Pisin). In example 194a, the object pronoun is explicitly stated as third plural *olketa*. The *(-)im* cannot be interpreted as a pronoun, suggesting it has decategorialised from a pronoun to an agreement marker of a number invariant third person object. Correspondingly, the source meaning and target meaning have become separated. Today, *-im* in Solomons Pijin has conventionalised as a transitiviser suffix.

(194) Switching Context

a. Mi no luk**-im** olketa.
1sg.ERG NEG see-TRNS 3pl
I didn't see them.

(Keesing 1991:319)

Theme markers in Papuan languages develop out of preferred discourse contexts (De Vries 2006:814-5, Reesink 1994:118), and the linguistic source materials are also identified as being similar. De Vries (2006) notes that Papuan languages that do not have a dedicated theme marker utilise deictics to mark theme constituents

(p.814). Thematising strategies in Papuan languages can also take the form of subordinate structures, which are often translated into English by adverbial and relative clauses (Foley 1986:201). The Nmbo phenomenon here also arises from these preferred contextual uses, as I will outline below.

The Nmbo DEM+COP construction stages go as follows. In the initial stage we have adjoined subordinate clause constructions of the type mentioned earlier (the same examples are provided below). The verb in the subordinate clause can also be a copula, which would agree in person/number and TAM with the relevant constituents of the complex clause (example 195c). Abstractions of the structural and semantic possibilities of the initial stage are shown in table 8.3

Structure Options	Meanings Available
[NP DEM_{LOC} V_{TAM} $]_A$ [(NP) V_{TAM} $]_B$	"The NP_A that $\text{V}_A$ed, it $\text{V}_B$ed."
[NP DEM_{LOC} COP_{TAM} $]_A$ [(NP) V_{TAM} $]_B$	"The NP_A that is here, it $\text{V}_B$ed."

Table 8.3: Structural and semantic possibilities of DEM+COP initial stage.
nphd = non-prehodiernal tense

(195) [Zoga is talking about the time an Australian man came to the area to recruit workers for a diving job. The sole argument of the subordinate clause is also A of the matrix clause verb.]

a. [yna site är markai **ge** **t-n/maro**]$_S$
DEM.prox white man foreigner DEM 3sg.nphd/come
[n-n\ovar/yng Sigabadru hamba=t.]$_M$
3sg.A:pfv.pret:ven/arrive Sigabadru village=ALL
[This white man that came] arrived at Sigabardu village.

G20150826-03AusWorkZG, 00:04:06.430- 00:04:15.545

(196) [Rachel is speculating about her new born daughter's future. The sole argument of the subordinate clause is also A of the matrix clause verb.]

a. [bä kitong **ge** **k\amdo/nga**...]$_S$ [elementrian

FUT big DEM 3sg.A:pfv.fut/become elementary.school

ä k\ilau/nga...]$_M$A

FUT 3sgA:pfv.fut/enter

When she becomes big she will enter elementary school (lit. [She that will become big,]$_S$ [(she) will enter into elementary school.]$_M$

WSEK1-A20151001-09PaulaRM, 00:00:16.655 - 00:00:18.106.

(197) [Gabriel is giving biographical information.]

a. [Ynd **ge** **w/m**]$_{REL}$ [Govav-n skool-an nilawoyn]$_{M}$

1ABS DEM 1sg.nphd/be Govav=LOC school=LOC 1sg.pfv.pret/enter

[Me,] I entered the school at Govav.

WSEK1-G20170630-01GW02MQhm, 00:09:11.182 - 00:09:13.300

Reanalysis from source to target occurs in the **bridging context**. The most frequent context in which the subordinate clause occurs is where the referent of the relative clause NP is third singular, and the main verb TAM is nphd. It has been proposed that grammaticalisation is triggered in the main by frequency of use and context (Bybee and Hopper 2001, Diessel 2007), and there is evidence from a corpus of Nen that the most frequently expressed person/number and TAM is the 3sg and ipfv.nphd (Muradoğlu 2017:25). Assuming this distribution in Nen is applicable to Nmbo (and we will see later that this is the case), we can ascribe this general frequency of the 3sg person/number and nphd TAM as the bridging context.

In this bridging context stage the copula in the relative clause will take the 3sg.nphd form *ym*. The interlocutor (and possibly the speaker) do not notice (or mind) whether the DEM+ym form is agreeing or not with the NP$_A$ or VP$_B$. It is in this invisible context where metonymic semantic reanalysis occurs, from a "here it is" to a presentative topic marker. We can hypothesise that this bridging context is also where phonetic reduction occurs most frequently, from disyllabic *ge.ym* or *g.ym* to monosyllabic *gym*. An abstraction of this Nmbo bridging context is shown in table 8.4.

Structure Options	Meanings Available
[NP DEM$_{LOC}$ *ym*]$_A$ [(NP) V$_{nphd}$]$_B$	"The NP$_A$ that is here, it V$_B$ed."
[NP DEM$_{LOC}$ *ym*]$_A$ [(NP) COP$_{nphd}$]$_B$	"The NP$_A$ that one, it V$_B$ed."
[NP **ge ym~gym**]$_A$ [(NP) V$_{nphd}$]$_B$	"The NP$_A$ that one, it's like so."

Table 8.4: Structural and semantic possibilities of DEM+COP bridging context. nphd = non-prehodiernal tense

(198) [Bridging context examples.]

a. [[tande ama]$_{3sg}$ **gym**,] [Nmbo zi n\owav/t$_{nphd}$.]

1sg.POSS mother gym Nmbo language 3sg.A:ipfv.nphd/speak

[As for my mother,] she speaks the Nmbo language.

WSEK1-G20170713-02MQMW

b. [[Zi]$_{3person}$ **gym**,] [yao kor tomba y/m]

story.ABS gym NEG really long 3sg.nphd/be

[As for the story,] it isn't long.

WSEK1-B20170621-01KS01Ypsthm

The switching context is where multiple interpretations become more difficult or impossible. The *gym* no longer agrees with NP$_A$ or V$_B$ and therefore cannot be understood as a copula. This decopularisation is only visible when NP$_A$ person/number is not 3SG, or when V$_B$ TAM is not NPHD.

Structure	Meaning
NP **gym** (NP) V$_{TAM}$	" The NP [**topic**] it V$_{TAM}$ed."

Table 8.5: Structural and semantic possibilities of DEM+COP switching context. TAM = any possible TAM specification

(199) a. yndo **gym**, tanzo är bérbér-uva t\ake/tawn.

1sg.ERG **TOPIC** 1sg.POSSC husband fear-COM 1sg.A>3sg.U:ipv.rmpst/watch

I [**TOPIC**] looked at my husband with fear.

WSEK1-B20150813-02HomHZ, 00:02:50.880 - 00:02:53.755

The example above is, however, technically still ambiguous in its interpretation. The NP *yndo* (1sg.ERG) is physically present at the time of the utterance, so an alternative interpretation of the clause, which is closer to the source meaning, could be "Me here I am, looked upon my husband with fear." Other examples include speakers gesturing to an NP referent that is physically present, and talking about how it was used in the past. In the following example, Taqm had

just finished a demonstration of an instrument that was used during initiation rituals in the past. The DEM+COP follows the NP "this custom" which is understood as the event which was just demonstrated. The main verb refers to the state of the past, when the custom was actually practised. Such examples are not taken as a problem in the analysis of this proposed grammaticalisation process. They merely highlight that 1) the bridging and switching contexts are not discrete stages but can have much overlap, and 2) there are many points of ambiguity in Nmbo topic constructions that can be used by speakers creatively; this merely widens the pool in which reanalysis can occur.

(200) [Taqm is giving an explanation of how the pipe instrument *ali* was used in initiation rites of the past.]

a.	yände	yam	**giym**,	dena	bä	t/maro.
	3sg.POSS	custom.ABS	**TOPIC**	DEM.mn	3ABS	3sg:rmpst/be

Its custom [**topic**] it was like so.

WSEK1-B20170615-01AriTS5s

8.4 Data, and Method of Investigation

The data for this study is mostly based on the Nmbo Sociolinguistic Corpus with a few additions from other recordings that were made for documentary purposes. There are 513 tokens of clauses with DEM+COP, where the NP is headed by a referential N, and the V is inflecting. Each clause was identified manually by inspecting the orthographic transcriptions and/or listening to recordings. Topic constructions where the topicalised NP is a headless or adverbial NP have been omitted from the dataset. We will investigate three parameters of grammaticalisation in order to characterise DEM+COP variability: person/number (PN) and/or TAM decategorialisation of the copula, phonetic reduction of *gym*, and semantic/discourse patterns of DEM+COP.

Morphosyntactic decategorialisation for the DEM+COP is measured by looking at the loss of agreement by the copula of the DEM+COP: PN decategorialisation, and Verb TAM decategorialisation. For both parameters, a decategorialisation score is assigned. The scores are a binary assignment of agreement between the

DEM+COP and NP_{PN}, and the DEM+COP and the VP_{TAM}. For each topic construction in the corpus, the degree of decategorialisation was measured.

Score	DN (PN Agreement)	DT (TAM Agreement)
1	NP_A is not 3sg, COP is not 3sg *yndo wm* (1sg.ERG 1sg/be)	VP_B is not nphd, COP is not nphd *wngmaro wmaro* (rmpst/go rmpst/be)
2	NP_A is not 3sg, COP is 3sg *yndo ym* (1sg.ERG 3sg/be)	VP_B is not nphd, COP is nphd *wngmaro ym* (rmpst/go nphd/be)
3	NP_A is 3sg, COP is 3sg *ymo ym* (3sg.ERG 3sg/be)	VP_B is nphd, COP is nphd *wngm ym* (nphd/go nphd/be)

Table 8.6: Decategorialisation scores.
1 = In agreement
2 = Decategorialisation,
3 = Invisible Context.
The Nmbo verb examples in the "TAM Agreement" column *wngm* and *wngmaro* are both in a first person singular form.

Decategorialisation of person/number (DN) and TAM (DT) are both scored between 3-0. A score of 3 signals an invisible context. A DN score of 3 means the NP_A and DEM+COP are both 3sg, and a DT score of 3 means the VP_B and DEM+COP are both nphd or a best match TAM. (see table 8.6 for score). These are the contexts where we cannot tell whether there is actual or incidental agreement occurring between the constituents. A DN score of 2 is an NP_A with a person/number that is *not* 3sg, but the DEM+COP is 3sg. In this case, we can see clearly that there is disagreement, i.e. decategorialisation between the DEM+COP and the NP. Likewise a DT score of 2 means the VP_B TAM is *not* nphd, but the DEM+COP TAM is, i.e. a case of decategorialisation between the DEM+COP and VP_B. A score of 1 means that the DEM+COP is not 3sg.nphd gym, but is in agreement with the NP_A and VP_B which are also not 3sg or nphd. A score of 0 was assigned for any mismatches that can only be explained as speech errors (of which there were only a handful). A summary of these scores and their correspondences are shown in table 8.6.

For example, example 201 has a 3sg NP_A, a nphd VP_B, and the DEM+COP is 3sg.nphd *gym*. In such a context we cannot tell if decategorialisation is happening. The decategorialisation score for this kind of clause is a 33, labelled as *INVIS*.

(201) *INVIS*: Invisible context].

a. **mrzu-m** **gym,** äuya **y\ere/t.**

girl-ERG$_{3sg}$ **TOPIC**$_{3sg.nphd}$ cassowary.ABS 3sg.A>3sg.U:**ipfv.nphd**/hold

The girl **[TOPIC]** she is holding the cassowary.

WSEK1-B20170615-02Taqm03Casshm

Example 202 has a 3sg NP$_A$, and the DEM+COP *gym* is 3sg which gives the person/number score a 3. But the VP$_B$ is in the pfv.pret, while the DEM+COP TAM is nphd. In this case we can see disagreement of TAM between the two constituents, so received a TAM score of 2. The total decategorialisation score for this clause is a 32, labelled as *INDT* (Invisible P/N Decategorialising TAM).

(202) [*INDT*: Invisible P/N, but decategorialising TAM example.]

a. Okei **mrzu-m** däv gymn, band tar-h

DISC **girl-ERG**$_{3sg}$ when **TOPIC**$_{3sg.nphd}$ ground.ABS dig-INF

y\avngo/y.

3sg.A>3sg.U:**pfv.pret**/do

Ok then the girl [TOPIC] she started to dig the ground.

WSEK1-B20150805-01NinyiTSPM

A final example 203 shows cases of perfect agreement where the DEM+COP is not *gym*. The person/number of NP$_A$ is 1sg, the TAM of VP$_B$ is rmpst, and the copula agrees with both taking the form *qmaro*. The clause score here is 11, labelled *ANAT* Agreeing Number, Agreeing TAM.

(203) [*ANAT*: Agreeing Number, and Agreeing TAM]

a. **ynd** nmbone **ge qmaro** yndo tanzo äusa

1ABS small TOPIC$_{1sg.rmpst}$ 1sg.ERG 1sg.POSSC old.lady.ABS

t\ake/tawn.

1sg.A>3sg.U:**ipfv.rmpst**/see

Little me [TOPIC] I was watching my grandmother.

WSEK1-G20151015-15BasketMWPM

A summary of the scores and their meanings are shown in table 8.7.

PN	TAM	Clause	Code	Interpretation
3	3	33	INVIS	Invisible Bridging Context
3	2	32	INDT	Invisible PN, Decategorialised TAM
3	1	31	INAT	Invisible PN, Agreeing TAM
3	0	30	INET	Invisible PN, Error TAM
2	3	23	DNIT	Decategorialised PN, Invisible TAM
2	2	22	DNDT	Decategorialised PN, Decategorialised TAM
2	1	21	DNAT	Decategorialised PN, Agree. TAM
2	0	20	DNET	Decategorialised PN, Error TAM
1	3	13	ANIT	Agreeing PN, Invisible TAM
1	2	12	ANDT	Agreeing PN, Decategorialised TAM
1	1	11	ANAT	Agreeing PN, Agreeing TAM
1	0	10	ANET	Agreeing PN, Error TAM

Table 8.7: Decategorialisation Scores for the topic clause, and the interpretation of the scores.

Phonetic reduction was measured by looking at syllable count, and duration of DEM+COPs in the form *gym*. For syllable count, only *gym*s were considered, i.e. comprised of the demonstrative in *ge* form (not *gs*), and the copula in the *ym* form (not *ymn*). For duration, only monosyllabic *gym* were considered. Duration was measured for each topic construction, as well as for the *gym* in the clause. Onset and offset were located by hand on the basis of a waveform in ELAN.

For the measures of phonetic reduction, only the disyllabic *ge ym* and monosyllabic *gym* were considered, i.e. tokens of *ge ymn*, *gs ym*, and *gs ymn* are excluded. Duration for monosyllabic *gym*s were calculated for each observation. Speech rate was controlled for in the following way. The number of syllables in each topic construction were counted (c_{syll}), then divided by the raw duration d (in milliseconds) of the same construction. The duration of the *gym* token was then divided by the construction duration. Finally, the log value of this duration was taken.

$$C = \frac{c_{syll}}{d_c}$$

$$G = log\frac{d_g}{C}$$

Where d_c is the duration of the construction, and d_g is the duration of the monosyllabic *gym* token. C is the controlled construction duration, and G is the speech rate controlled *gym* duration.

For semantic/discourse characterisations I used a modification of the Givón (1983) method for tracking topic continuity. Givón traces topic continuity by three measurements: look-back/referential distance, ambiguity/potential interference, and persistence/decay. Of these measures, I applied *Look-back* to the DNDT constructions to see whether there is a tendency for one kind of usage,[2] in the cases where we can see decategorialisation occurring for P/N and TAM. *Look-back* measures the gap between previous occurrence of a referent/topic in discourse. In practical terms, this means counting back the number of clauses from a topic NP to the last prior mention, as implemented by Hinds (1983).

8.5 Results: Evidence of Grammaticalisation

There is evidence of DEM+COP morphosyntactic decategorialisation occurring at low frequencies in Nmbo. Of the 513 tokens collected, 338 (65.9%) are the morphologically 3sg.nphd *gym*.[3] 113 out of the 338 total *gym* occurrences show either PN or TAM decategorialisation (33.4%). Cases where we can see both P/N and TAM decategorialising are only 4% (n = 15/338) of the total number

[2]I attempted to apply ambiguity/potential and persistence/decay to the DNDT *gym*s, but the scores mostly overlapped with *Look-back* so will not be discussed in addition.

[3]This count includes alternate forms such as *gs ym*, and *ge ymn*.

of *gym*s, which is 3% of the total dataset (n = 15/513). While low, these numbers suggest there are small amounts of decategorialisation occurring.

Most *gym* tokens are occurring in the invisible bridging context. Of the morphological *gym*s, the majority occur in an environment where both the NP is 3sg, and V is nphd or a possible best fit TAM (n = 225/338, 67%). These invisible *gym*s make it impossible to tell whether these are cases of genuine agreement or the use of a grammaticalised *gym* which is incidentally in a 3sg and nphd context.

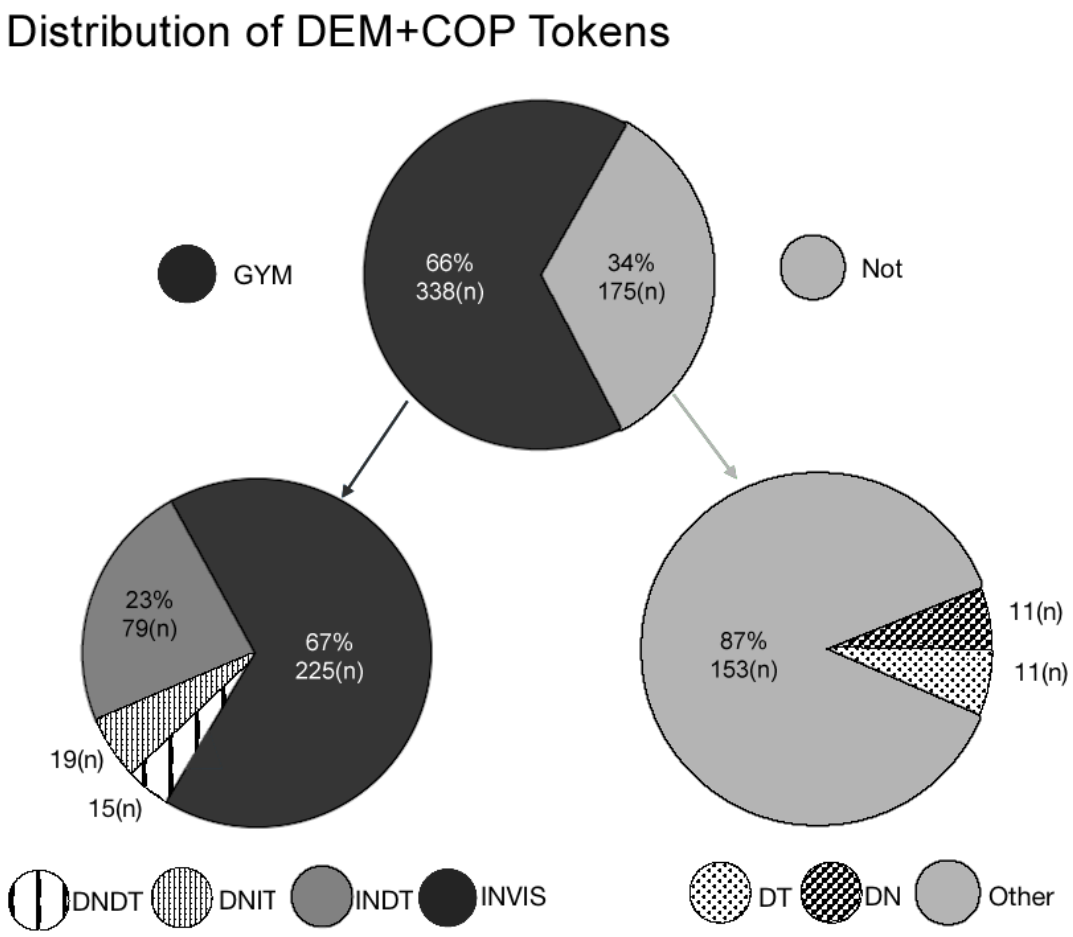

Figure 8.4: Distribution of DEM+COP tokens. DNDT = Decategorialised PN, Decategorialised TAM; DNIT = Decategorialised PN, Invisible TAM; INDT = Invisible PN, Decategorialised TAM; INVIS = Invisible Bridging Context; DT = Decategorialised TAM; DN = Decategorialised PN; Other =

There is evidence that 3sg.nphd *gym* is the resultant form of decategorialisation, and not some other DEM+COP form such as *ge tmaro* (3sg.rmpst) or *ge wm* (1sg.nphd). Of the 175 DEM+COPs that are not *gym*, there are only two example where both P/N and TAM are both decategorialising: one each for *ge ymaro*

(3sg.prim) and *ge tm* (3sg.ypst) (table 8.8). We may expect to see similar levels of decategorialisation for other DEM+COP forms if the *gym* rates are reflecting a general tendency and not something unique to *gym*. This is, however, not the case. The three most common non-*gym* DEM+COPs are *ge tmaro* (3sg.rmpst, n = 58), *ge ymaro* (3sg.prim, n = 32), and *ge wm* (1sg.nphd, n = 26). If, say, *ge tmaro* were also a decategorialising form, proportionally we may expect around 27.8% TAM decategorialisation, and/or 10% of P/N decategorialisation in line with rates for *gym*. The form *ge tmaro* does not show this pattern (TAM decategorialisation = 8.6%, P/N = 6.9% see table 8.8). P/N decategoriaisation for *ge ymaro* is higher than *gym* (12.5% vs *gym*'s 10%), while TAM decategorialisation for *ge wm* is somewhat close to *gym* (23.1% vs gym's 28%). We may find more cases of non-*gym* DEM+COPs decategorialising if the data set were expanded, but with this current set it is clearly *gym* that is showing overall highest rates for decategorialisation in both inflectional categories of TAM and P/N.

DEM+COP	Decategorialisation Labels								Decategorialised	
	INAT	DNAT	DNDT	INDT	ANAT	ANIT	ANDT	Total	TAM	PN
ge tmaro	49	4		5				58	8.6%	6.9%
ge ymaro	25	3	1	3				32	12.5%	12.5%
ge wm					1	19	6	26	23.1%	0.0%
ge qmaro					12	0	2	14	14.3%	0.0%
ge tm	7	1	1	1				10	20.0%	20.0%

Table 8.8: Distribution of non-*gym* DEM+COPs.
ge tmaro = 3sg.rmpst; *ge ymaro* = 3sg.prim
ge wm = 1sg.nphd; *ge qmaro* = 1sg.rmpst; *ge tm* = 1sg.ypst

Decategorialisation of *gym* appears to be progressing by a two-step process where TAM decategorialises faster than PN. 27.8% of the *gym*s show some TAM decategorialisation (n = 94), while PN decategorialisation is at 10% (n = 34). For the non-*gym*s, TAM decategorialisation makes up 45.1% of cases (n = 79) while PN decategorialisation stands at 7% (n = 12). In fact, the decategorialisation of TAM appears to also be more common for the non-*gym* DEM+COPs (table 8.8).

The prevalence of TAM decategorialisation over that of P/N decategorialisation hints at the semantic origins of DEM+COP as a grammaticalising topic marker. The DEM+COP construction is such that the topicalised NP and the V can occupy different temporal locales in the telling. Example 204 has the speaker topicalising "Masa's wish", but flagging that it existed in the past by the use of the rmpst form of the copula *tmaro*. The desire then turned into something, for which

the speaker uses the pfv.fut tense. Moving between TAM specifications within a clause is a cross-linguistically common strategy for creating narrative effects (e.g. Fleischman 1989 for English, Stirling 2012 for Kala Lagaw Ya (Pama-Nyungan; Australia), Martin 2010 for Creek (Muskogean; USA)).

(204) [Gima is talking about the reason why a particular tree was planted at a particular location.]

a. Masa mende **ge** **tmaro**, bä biträr=ot
Masa wish DEM 3sg.**rmpst**/be 3ABS tree.type=ALL
k\owano/ngai.
3sg.A:**pfv.fut**/become
Masa's wish [(which was in the past) TOPIC] turned into a biträri tree.

WSEK1-G20170622-01GimaGZ-02MQhm

For the parameter of phonetic reduction, there is no evidence that degree of decategorialising has an effect on the duration of *gym*s. If phonetic reduction is a gradual process, we may see evidence of this gradation in the duration data. The durations of the *gym*s, however, show no meaningful differentiation. The mean durations of the three decategorialised *gym*s is 10.62, and the mean of the invisible *gym*s is 10.6. Since reduction is uniform across the decategorialised types of *gym*s, it seems that phonetic reduction has either been been completed, or the reduced phonetic context was part of the initial context. Bybee (2012, 2017) argues that phonetic conditions and reductions are what triggers decategorialisation in the first instance, and the uniform distribution of *gym* durations could be interpreted as such.

While the token counts are very low, there is some suggestion that the fully decategorialised *gym* clauses have discourse functions that are more expanded than what is generally claimed for subordinate clauses (i.e. the initial context). While relative clauses delimit the range of possible referents (Andrews 2007:207), decategorialised all instances of decategorialised *gym* resulted in a *Look-back* score of 1 or 0; that is, the topicalised NP is mentioned in the immediately preceding clause (1) , or is introduced as a new referent by *gym* (0). Exactly half of the DNDT *gym*s had a score of 1, and the other half 0 (n = 6/12). When the

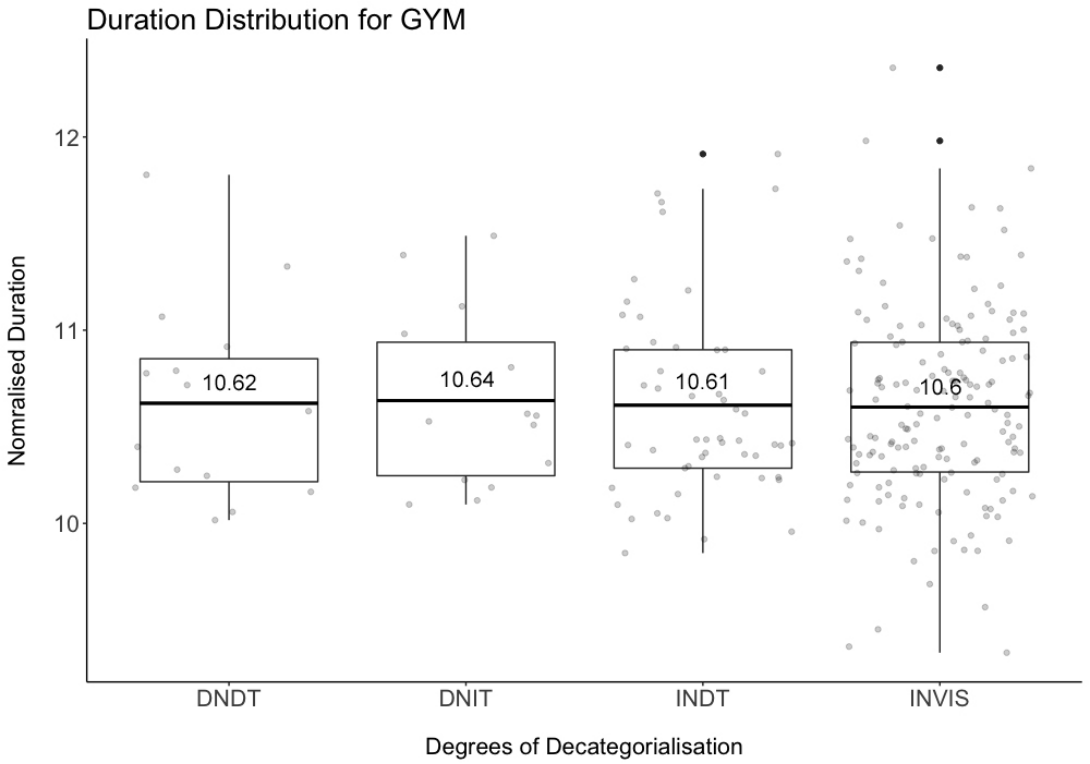

Figure 8.5: Duration of *gym*s grouped by decategorialisation type. The central black bar in the box plots show the mean. The dots represent each occurrence of *gym*.

Look-back score is 0, the topicalised NP is either introduced as a new referent, or as in the case of example 8.5a, limits the agentive NP to the 1sg from what was the 1du topic in the preceding clauses. When the *Look-back* score is 1, the topicalised NP is a re-emphasis of the referent which was just mentioned in the preceding clause (example 8.5b).

[In the following examples, the numbers on the right trace the number of clauses back from the topic construction under investigation. The target clause is numbered 0, so a clause numbered 1 means it is 1 clause prior to the target constriction.]

Emphasis shift (8.5a)

5. *tnrwn bavwa...*
I turned over the taros...

4. *nne...*
the yams...

3. *se kotarwm,*
the two of us called out,

2. *nnowaneaem,*
the two of us arrive,

1. *ynḡn ynvlmm...*
we two put them in the bags...

0. ***yndo gym tanzo ynḡ... yilawoyn,***
I [TOPIC] put my bag inside,

A. *deve Tokande zärngun...*
father Toka's yam house...

Re-emphasis (8.5) [Speaker TS is describing the time he was fighting off a boar that was attacking him.]

4. *mamwi kt kotnawng tokn,*
The boar was shouting above [of me],

3. *"Wä~!*

2. *Mamwi wlqmae yrmdräryn,*
I was still holding the boar tightly,

1. *mamwi muyavna namdoyng,*
The boar became without strength,

0. yände muye mnutyu *giym* nmbnarvi namdoi.
All of his strength [TOPIC] became small.

Another interesting pattern pertains to the grammatical and semantic role of the topicalised NP. Nmbo, with its ergative/absolutive case system, shows patterns consistent with the observation that semantically agentive NPs are are more likely to undergo topic marking. As Du Bois (1987) noted, new information in discourse appears preferentially in S or O roles, while topic tends to be an NP which is underlyingly S or A in function. Fox and Thompson (1990) found English relative clauses with subject heads tend to be characterising clauses, as opposed to object head relative clauses which convey new information (pp.306-307). Bickel (2000) notes how ergative clauses primarily signal sentential topic in Kiranti languages (Sino-Tibetan; Nepal and India).

Ergative marked NPs account for 23.4% of the entire *gym* dataset (n = 79/338), but the occurrence is much higher in a subset of decategorialisation types. We find that PN decategorialising clauses (DNDT and DNIT) show an ergative

marked topicalised NP over 70% of the time (n = 24/34). On the other hand *gym* clauses where TAM is decategorialising (DNDT and INDT), the topicalised NP is ergative only 18.7% of the time (n= 57/304). In the invisible context it is less common to have the topicalised NP in ergative, with only 15.1% of the data showing this pattern (n = 34/225). This suggests that the more agentive the NP, the more likely it is topicalised, and in such cases, there is a high chance that the DEM+COP PN will be decategorialised to a 3sg form.

	ERG NP		ABS NP		
	n	%	n	%	Total
DNDT	9	60.0%	6	40.0%	15
DNIT	13	68.4%	6	31.6%	19
INDT	23	29.1%	56	70.9%	79
INVIS	34	15.1%	191	84.9%	225

Table 8.9: Token count of NP semantic role in *gym* constructions by subset of decategorialisation type.

The Nmbo data on discourse and semantics suggests that there is a preferred semantic/discourse usage of decategorialised *gym*s. These are the agentivity of the topicalised NP, and the function of introducing a new NP into the discourse, but also to shift emphasis and re-emphasise the NP. With an expanded dataset this finding can be made more robust. Given the breadth of research that covers agentivity, ergativity, and topicality, further discussions pertaining to the tantalising findings of Nmbo must be parked for future research. The interim conclusion is that the preferred semantic/discourse usage of the topicalising *gym* may be as an emphasiser of agentive arguments.

In summary of this section, we find evidence that *gym* is the target form of DEM+COP decategorialisation. Not only is DEM+COP decategorialising as *gym*, different parameters are grammaticalising at different rates. TAM decategorialisation is more frequent than P/N decategorialisation. As far as phonetic reduction is concerned, decategorialised *gym*s have similar durations to agreeing and invisible *gym*s. There is evidence of an emergent semantic/discourse preference for these *gym*s as emphasisers of agentive NPs.

8.6 Results: Sociolinguistics of GYM

With only 113 of *gym*s showing some kind of decategorialisation, we cannot detect any social patterns to DEM+COP decategorialisation.

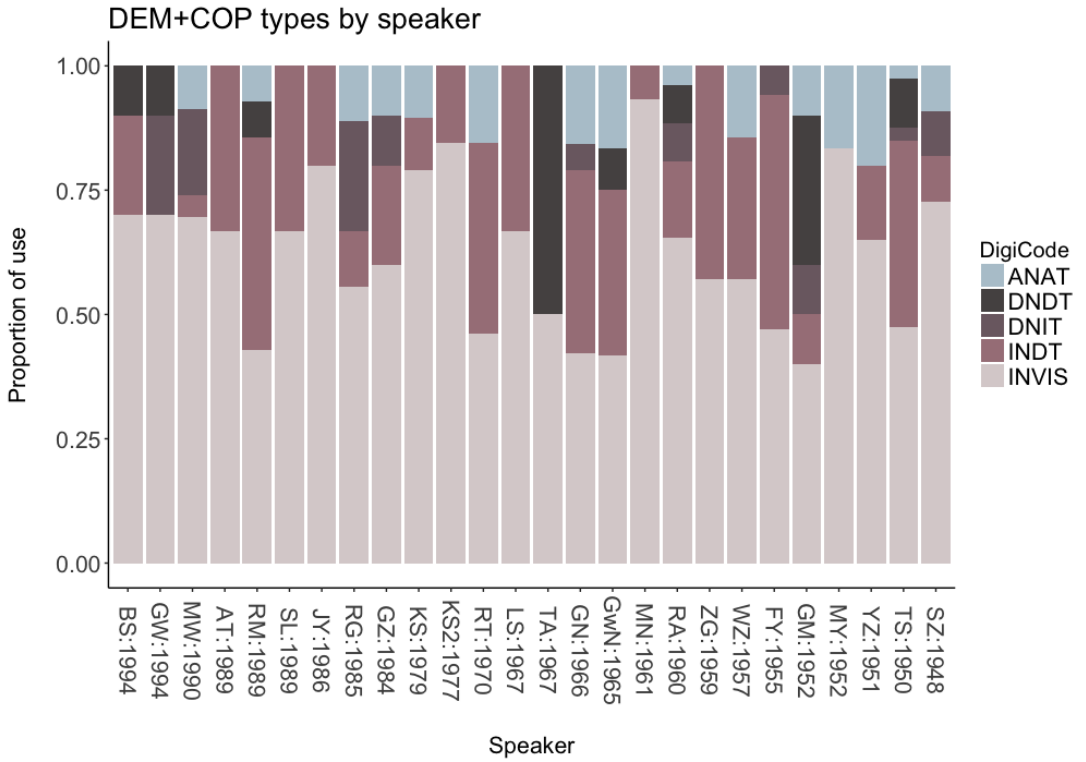

Figure 8.6: Proportional representation DEM+COP types by speaker, organised from youngest to oldest. Year of birth follows the initial of the speaker.

The bar graph 8.6 shows the rates and proportions of decategorialised *gym*s organised by speaker. The speakers are arranged by year of birth on the x-axis, with the younger speakers on the left. The impression one gets is that, as far as rates are concerned, the speakers are quite uniform. The numerical values represented in the bar graph is presented in table 8.10. The average proportion of decategorialised *gym* per speaker is 32.3%. There are, however, a handful of individuals with relatively high token counts showing decategorialisation proportions of nearly 50% (speakers FY:1955, TS:1950, GM:1952, TA:1967, RM:1989), but it is unclear how representative this is of these individuals' tendencies, or the speech community more broadly. Speaker TS:1950 with 50% of his 40 tokens showing some kind of decategorialisation is quite a robust figure, and speaker FY:1955 is also shows a high 52.9% of decategorialising *gym*s, although her total token count is much lower than TS:1950 at 17. Curiously both individuals are on the older spectrum of speaker age, suggesting that there is no evidence

of a correlation between age and *gym* decategorialisation. Pearson's correlation returns a figure of -0.15, meaning there is little evidence of an age effect when it comes to the use of decategorialised *gym*.

If we limit out numerical observations to speakers with a total tokens count over 19 we can make some conjectures as to what picture additional data may paint. The average proportion of decategorialised *gym* for these 6 speakers is 28.3%, which is slightly lower than the overall sample population. A Pearson's correlation on this group, however, returns a weak to moderate result of -0.44, suggesting that the older the speaker, the more likely they are to have a higher proportion of *gym*s that are decategorialised. This is a slightly counter-intuitive result which is likely affected by the skew in the sample, but given that the total sample population's correlation was also negative, this may be taken as some weak evidence that older speakers are behaving somewhat differently to other speakers.

An attempt was made to test Village as a variable, since this social variable came up as significant in both the vowel sociophonetic study, and the [h]-drop study. Token counts for decategorialised *gym*s were too low to test directly, so Bayes factors were calculated[4] to test the hypothesis of whether different groups have similar underlying probabilities of producing INVISible *gym*s, or not. If the production rates of INVIS are different across villages, we can take that as some evidence of differential behaviour across the two villages; ergo, some evidence that rates of non-INVIS *gym*s are different.Two conditions were tested: speakers grouped as Originating from Bevdvn vs Not, and Residing in Bevdvn vs. Not. The origin village condition returned a Bayes factor of 20.1, while the Residence village condition returned 19.3, in favour of the distributions being the same. Both provide strong evidence that Bevdvn speakers have the same underlying probability of producing INVISible *gym*s as speakers from other villages; in other words, that the Nmbo speakers in the dataset all produce INVISible *gym*s at similar rates. The implication is that, as far as decategorialised *gym*s are concerned, Nmbo speakers are part of the same speech community with similar rates and norms.

The lack of patterning by social variables such as age or village suggests that *gym* decategorialisation is in stable variation. In theory, additional data with an

[4]These calculations were made by T.M. Ellison in context of Evans, Ellison, and Kashima (2018)

increase in token counts per speaker may reveal some evidence of social diffusion. The suggestion that older speakers may be using decategorialised forms more frequently could be made more robust with additional data. There is, however, no reason to assume that there is any social diffusion of DEM+COP decategorialisation, since linguistic variants can remain stable within a linguistic system for many decades, and even centuries (Mair 2004 for examples in English).

8.7 Discussion and Conclusion

I have described the morpho-syntactic decategorialisation of DEM+COP as an emergent topic marker, where TAM decategorialisation is progressing further along than P/N decategorialisation. The two-step process may be due to creative tense-shifting for narrative effect, where the topicalised NP and the main verb V show a discrepancy in temporality. There is also some evidence that topic clauses with ergative marked NPs have a higher rate of decategorialised *gym*. Phonetic reduction was investigated, but the duration of decategorialised *gym*s shows no difference to *gym*s in non-decategorialised contexts. The discourse function of decategorialised *gym*s appear to be that of re-emphasising referents that have just been mentioned, or to shift emphasis on to a referent that was mentioned in the preceding segment. All of this leads DEM+COP variation to be characterisable as an instance of grammaticalisation[5], or pragmaticalisation, depending on one's definition of grammar and pragmatics (c.f. Degand and Evers-Vermeul 2015).

The sociolinguistic patterns are not clearly interpretable. There is some weak evidence that older speakers are the ones to use decategorialised *gym*s, but the more straight forward interpretation is that gym decategorialisation is in a stable

[5]An examiner has suggested that gym could be analysed as a focal topic marker, in contrast to *ge* and *gs* which already seem to be functioning as general topic markers. The suggestion was that the copula component of *gym* provides the semantics of focality, as opposed to *ge* and *gs* which are more general thematising markers. While this is a possible outcome, I have chosen not to speculate on the possible semantic outcomes of grammaticalising gym. The Nmbo results of the discourse and semantic analysis neither support nor deny the possibility of a focal gym vs general ge/gs contrast. If *gym* were to be contrastive with ge and gs at this current stage, we may want to find constructions such as "yna rokar **gym tmaro**, kakiyam tavrotao", where *gym* is completely substitutable with *ge*/*gs* in subordinate constructions. In such cases *gym* could occur preceding a fully inflecting copula (*tmaro* in the example just given). I do not find such constructions in the current state of the corpus. Since I do not have any evidence of this possible outcome, and I do not wish to speculate about possible outcomes beyond what we can see synchronically, I have chosen not to incorporate into the main text. It is, nonetheless, an interesting suggestion, hence I have noted it here in a footnote for future reference.

Speakers	Agr (n)	Prop.	Decat (n)	Prop	INVIS (n)	Prop	Total (n)
BS:1994	0		3	30.0%	7	70.0%	10
GW:1994	0		3	30.0%	7	70.0%	10
MW:1990	2		5	21.7%	16	69.6%	23
AT:1989	0		1	33.3%	2	66.7%	3
RM:1989	1		7	50.0%	6	42.9%	14
SL:1989	0		1	33.3%	2	66.7%	3
JY:1986	0		3	20.0%	12	80.0%	15
RG:1985	1	11.1%	3	33.3%	5	55.6%	9
GZ:1984	1	10.0%	3	30.0%	6	60.0%	10
KS:1979	2	10.5%	2	10.5%	15	78.9%	19
KS2:1977	0		2	15.4%	11	84.6%	13
RT:1970	2	15.4%	5	38.5%	6	46.2%	13
LS:1967	0		2	33.3%	4	66.7%	6
TA:1967	0		1	50.0%	1	50.0%	2
GN:1966	3	15.8%	8	42.1%	8	42.1%	19
GwN:1965	2	16.7%	5	41.7%	5	41.7%	12
MN:1961	0		1	6.7%	14	93.3%	15
RA:1960	1	3.8%	8	30.8%	17	65.4%	26
ZG:1959	0		3	42.9%	4	57.1%	7
WZ:1957	1	14.3%	2	28.6%	4	57.1%	7
FY:1955	0		9	52.9%	8	47.1%	17
GM:1952	1	10.0%	5	50.0%	4	40.0%	10
MY:1952	1	16.7%	0	16.7%	5	83.3%	6
YZ:1951	4	20.0%	3	15.0%	13	65.0%	20
TS:1950	1	2.5%	20	50.0%	19	47.5%	40
SZ:1948	1	9.1%	2	18.2%	8	72.7%	11

Table 8.10: Token count and proportion of gym type by agreement, some level of decategorialisation, and invisible context

state of variation within the community. There is also no evidence of differential rates across villages.

In order to investigate topicalisation of ergative NPs, and the sociolinguistics of *gym* decategorialisation, the Nmbo data can be expanded by including data from neighbouring sister languages. The neighbouring Nambu branch languages all have cognate copular verbs with similar clausal syntax. An expanded dataset including neighbouring villages, who arguably constitute a part of a larger Nambu branch Morehead area speech community, would also help determine whether gym variation is stable, or has differential rates and parameters of change across other locales.

Part IV

Conclusion

Chapter 9

Conclusion

Ge ynamdamn, yakemtamn yna yu, namnamn, kkv yna bä yavrotan. "I examined it, I looked at this place, I thought, I will make a garden here." (Bévék Bévék, Short Story 2015)

One day during my final field trip in July 2017, there was another disturbance in the village of Bevdvn. A young man, known as a village hot-head who takes after his father in fiery temperament, attacked one of my consultants with an axe. My consultant was unharmed, but he told me of this encounter while his wife, and the daughter of his adoptive father watched on and nodded their support. "We're getting tired of that corner [that family] and their selfishness," my consultant had said. "We're thinking that we'll maybe stay at our garden place. The phone tower [in Bevdvn], we don't need these things. A few months, maybe a few years, we'll stay in our garden place." At the time of writing this thesis, my consultant had not acted on his words. His sentiment, however, left me wondering whether this is how social fissures and language diversification occur in the Morehead area; by physical relocation and avoidance of neighbours.

My consultant's words are not so unusual in the Nmbo speech community. It is a shame that I could not record this particular conversation, but I have a handful of other recordings where men talk about their intentions of leaving the village and relocating to new garden places. There are linguistic expressions that

tantalisingly seem to reflect the common occurrence of relocation and settlement due to social tensions. The Nmbo dictionary has multiple definitions for the verb *älmänh*, where one of the meanings is 'to be in disharmony with something or someone', and another 'to settle in a place, i.e. to make a settlement'. The desire to be free of social disharmony is not unique to the Morehead area, although it is my contention that the people here are practically capable and culturally equipped to do so should they desire.

In the qualitative part of this thesis I described Morehead area societies and the Nmbo speech community as an egalitarian multilingual society with a streak of individualism and tendency for out-group avoidance. The quantitative chapters suggest that village is a relevant social variable that explains patterns of micro-variation within the speech of Kerake Nmbo speakers. In this final chapter I will contextualise the qualitative descriptions with the findings of the quantitative studies in terms of our earlier discussions concerning the sociolinguistics of language diversity in Papua New Guinea.

9.1 Summary of the qualitative studies

In the introduction to this thesis I summarised three speech community parameters thought to be relevant to language diversification in New Guinea (section 1.5). I will address these parameters as a way of summarising the socio-cultural descriptions of the Nmbo speech community and the Morehead area.

1. Community configurations of a small population, coupled with dense multiplex networks which themselves are loosely connected over geographically contiguous areas.

The communities of the Morehead area are quite small in terms of absolute population size (chapter 2). My own fieldwork in the Nmbo speech community adds that the social units of interaction are also small, based on households as units of primary production (2.2), connected by the social exchange of marriage to other household and small groups (3). Social networks are thus inferred as dense and multiplex, but dispersed due to the tendency of individual autonomy and outsider avoidance (2.4). Having said so, the speech community of Nmbo is justifiably at the village level, built up by various group practices (3.1) and ideology. Villages are socially relevant places to which Nmbo is intrinsically related (2.6).

2. Socio-cultural pressures which are entailed in language ideologies of language differentiation, and more explicitly, language as emblematic of group identity.

Communities of the Morehead area operate within ideologies of language differentiation and emblemacity. Place and language are concepts essentialised in discourse, and utilised in the construction of individual and group identity (2.6).

3. Language ecology with multiple languages of relatively equal social power cultivates language ideologies of egalitarian multilingualism, which creates and maintains receptive-reciprocal multilingual interactions between speakers and groups.

Southern New Guinea is indeed an area with multiple languages, and as research in the Morehead area progress we see a complex tapestry of multilingual communities. Individuals have varying linguistic biographies and repertoires in terms of production and comprehension (2.7), and different villages seem to have different degrees of multilingualism (3.2). This study has presented empirical details of villages and some individual speakers, which suggest that marriage and an ideology of egalitarian multilingualism both sustain this linguistically diverse language ecology.

9.2 Summary of the quantitative studies

The patterns of linguistic variation studied in this thesis suggest that, as far as the variables studied are concerned, Nmbo speakers pattern much like one another. When small variations are found, however, these are based on village, i.e. geographically separate groups.

9.2.1 Vowel Sociophonetics

The Nmbo vowel space exhibits some systematic micro-variations, but speaker and word are highly significant variables that explain the pattern of variation. This is likely a partial reflection of the word list format the data were collected in. Despite this, we see evidence that speaker age, and village of origin come up as significant variables in explaining vowel height and fronting patters. Younger speakers are systematically lowering [u, e, o], and fronting [o, ə]. Govav villagers have a systematically lower realisation of the short vowels [ə, ɐ]. There is also evidence to suggest that Govav villagers, on average, have smaller vowel space

spread. Unfortunately not much can be said about vowel duration, except that there is some evidence that Bevdvn villagers show longer durations for the short vowel [ɐ] than Govav villagers.

Gender was significant a few times for some vowels and only in some measures. Gender was significant fewer times, and less systematically than age and village. There are also different gender patters across the two villages. Bevdvn male speakers realise [æ, ɐ] higher than female speakers, while this gender effect is absent among Govav speakers. Govav male speakers have a more backed [o] than Govav female speakers, but this effect is absent among Bevdvn speakers. Vowel space spread was also not found to be significantly different across genders.

9.2.2 Word initial [h]-drop

The word initial [h]-drop study shows a change in progress where speaker age is the single most effective variable that explains propensity for [h]-drop. The younger the speaker, the more likely that they near-categorically produce [h]-less words. The turning point in the speech community seems to be speakers born in the early to mid-1970s, where we can observe a drastic change in rates of [h]-drop. Like the vowel sociophonetic study, gender did not come up as a significant effect.

There is some evidence that village was a significant social factor in the past, but currently not so. Older Govav speakers show higher rates of [h]-drop compared to their peers from Bevdvn, but this village distinction is absent in younger speakers. We cannot tell if it is village of residence or origin that has the effect. The rate of [h]-drop across age groups is monotonic for the Govav sample population, but for the Bevdvn sample there is a more radical increase in rates of [h]-drop between the age groups, specifically between those speakers born before and after 1975. The interpretation of this data is that the speech community showing the monotonic rate of change is the origin of the [h]-drop phenomenon; Govav is the origin of [h]-drop, and Bevdvn villagers have adopted it rapidly.

The ethnographic and demographic details of Govav support the interpretation that it is the origin of [h]-drop. Govav is the larger village with a linguistically more diverse population than Bevdvn. The feature pool theory (Mufwene 2001, Cheshire et al. 2011) would predict that a population with more diverse linguistic inputs from other languages are a likely locus of linguistic innovation. Govav is the village where 76% percent of all household comprise multilingual

couples, while in Bevdvn this is only 20% (3.2). The story is not a straightforward one of licensed Nmbo speakers adopting [h]-drop from the multilingual feature pool since none of the major non-Nmbo languages spoken by in-marrying women have cognate [h]-dropped words. The [h]-drop may be just a typical and natural language internal process, but it is also possible that multilingualism in Govav generally encourages the unconscious adoption of innovations as a means of linguistic differentiation.

9.2.3 Emergent Topic Marker

The DEM+COP decategorialisation chapter investigated the emergence of *gym* as a topic marker in a mono-clausal complex, arising from a bi-clausal subordinate clause construction. The study investigated compression of phonetics, and morpho-syntactic structures, as well as semantic/discourse contexts. This makes the DEM+COP variable the one non-phonological variable investigated for this thesis, and furthermore one which spans many levels of linguistic structure. DEM+COP was investigated with a contextual model of grammaticalisation.

DEM+COP is decategorialising into a phonetically reduced form *gym*. The cases of unambiguously decategorialisated *gym*s are low in the overall dataset. As one would expect from a grammaticalisation process, a large portion of the data is in an invisible bridging context where we cannot detect whether the *gym* form is incidentally in agreement with the 3sg NP and NPHD or equivalent V TAM. From the linguistically detectable decategorialised *gym*s we find there is more TAM decategorialisation than of person/number. There is some evidence that the decategorialising *gym*s are also found to be topicalising ergative marked NPs.

There is no evidence that degree of phonetic reduction is associated with the degree of decategorialisation. The results show that the durations of decategorialised *gym*s follow the duration distributions of non-decategorialised *gym*s. This suggests that phonetic reduction is not necessarily a gradual process that runs alongside morphosyntactic decategorialisation. Phonetic reduction may proceed much earlier and faster than other forms of compression.

Sociolinguistic variables could not be identified from the DEM+COP data. The number of observations of *gym* is too low per speaker to say anything definitive about social diffusion. Older speakers and younger speakers seem to be producing

comparable amounts of decategorialised *gym*, but there is some weak evidence that perhaps older speakers are using decategorialised forms more frequently than younger speakers. The hypotheses are that this is either stable variation in the speech community, or change that is occurring so gradually that we would need a much larger dataset to identify the sociolinguistic characteristic of the speakers who are participating in, and leading, the grammaticalisation.

9.3 Discussion

The qualitative description of the Nmbo speech community fits the suggestions in the literature pertaining to the socio-cultural selection pressures of language diversification. The quantitative data do not clearly show a tendency towards differentiation. The minor relevance of place and the non-relevance of gender in explaining the patterns of variation in Nmbo suggest that the speech community does not have high levels of community internal variation. The Kerake tribe Nmbo speakers are patterning quite similarly to one another, including the women who have married out of their villages of origin. These results in turn raise questions about the literature discussed at the beginning of this thesis. In this section I will address some of these.

9.3.1 The relationship between micro- and macro- diversity

Macro-variations, such as global distributions of linguistic diversity, are understood as reflections of cumulative micro-variations (section 1.1). Speaker-centred models of language evolution such as the Utterance Selection Model (Croft 2000, 2006, 2008) offer a schematic that articulates the components involved in long-term change (1.2), but the model makes no predictions about the nature of variation within individual speech communities. One assumption that comes from the evolutionary sciences more generally is that micro- and macro-variation are in a scalar relationship; that large scale diversity reflects high levels of micro-diversity on the ground (Gillespie 1998:5).

We do not find high levels of internal variation in Kerake Nmbo, as far as we can tell from the variables studied. In fact, the results of the vowel sociophonetics and [h]-drop show high levels of similarity in terms of patterning. The similarity of speech and accents within Kerake Nmbo fits suggestions like those

made by Bowern 2010 and Trudgill 2011 that small-groups have a high degree of accent similarity. The DEM+COP study also show no clear sociolinguistic conditioning.

There are a few possible ways to interpret the similarities of patterning we find in the Nmbo data. One set of possibilities pertain to methodology. 1) The linguistic variables chosen for this thesis are perhaps not the variables that demonstrate high levels of variation. For example I mentioned in the sociophonetics chapter (6.4) that consonants would likely yield more variation than vowels do. A future study could compare the variability found in consonants with those found in vowels to better assess how much sociophonetic variability is to be found in Kerake Nmbo. 2) Another possibility for the similarity in the speech community concerns sampling style. The composition of the speech community may be such that the sampling method based on social groupings does not capture the speakers and variables that show great variation (Eckert 1989, although c.f. Walker and Meyerhoff 2013 for evidence of individual variability reflecting population level variability). A deep look at dialogue and interaction of individuals in the vein of Schilling-Estes (2004), the "short-thick" analyses of Meyerhoff and Walker (2012) may help to shed light on how to interpret the relative lack of variation we see in this project.

If the stability and similarities of patterning in the Kerake Nmbo data are not due to methodology, it may mean that variation within small-scale communities of linguistically diverse areas is not inherently high. There is no clear linguistic evidence that New Guinean speech communities have high levels of internal variability, or that they may have higher tempos of change.

9.3.2 Gender identity is subsumed under tribal identity

The quantitative studies suggest that, at least for the variables investigated, gender is not a tribe-internal linguistic community of practice despite being a cultural one (section 3.3.2). Village on the other hand is a cultural *and* linguistic community of practice. There are qualitative similarities and differences between these two communities of practice. In terms of similarities, both gender and village as communities are based on sustained face-to-face interaction in daily life. Daily life activities are both mix-gendered (e,g. subsistence work, 2.2), and segregated (e.g. cooking, hunting, 3.3.2). Given that both communities of practice share this characteristic of sustained face-to-face contact, mechanistic

explanations such as frequency of interaction seem unlikely to be a major factor responsible for the emergence of a linguistic community of practice.

The qualitative differences between the two communities of practice, then, are perhaps the more likely reason for why village has linguistic differences and gender does not. The importance of *place* as a local concept of identity is extended to the framing of intra-tribal linguistic difference. The propensity of speaking about other languages in terms of place are mirrored in the way Nmbo speakers identify Govav vs Bevdvn/Arovwe lexical differences. Recall also that Yarne Nambo/Namna, despite its linguistic similarities with Kerake Nmbo, is locally identified as distinct as the language of the Yarne tribe of the places Drdr and Pongarki (3.4). In other words, the notion of different people from different places as speaking slightly differently, is an idea that is extended to describe place-based variation within the Nmbo speech community.

I have suggested in the sociophonetic study that gender is linguistically non-relevant in the Nmbo speech community[1] due to the nature of cultural orientation in the Morehead area (6.4), where orienting against a linguistic and tribal Other is significant (2.6). The key point is that the act of orienting against other tribes subsumes tribe-internal gender differences. Linguistically marking gender is not a high priority. This is not to say, however, cultural orientation against an Other determines whether a speech community will show gender differences or not. Small-scale communities can show community internal gender differences, such as the Yami of Taiwan (population 300). Yami women were found to be raising the diphthongs *ay* and *aw* as a way of linguistically accessing the positive associations indexed on the phone (Rau et al., 2009). Minority groups with larger populations that orient against other groups can also still exhibit gender differences. In such cases gender distinctions are analysed as a proxy for variable social networks across genders. K'iche' men avoiding stigmatised form of fricated intervocalic /l/ is analysed as reflecting different social networks where men have more awareness that this form is stigmatised (Romero, 2009). Variables like gender are expressed in "highly local terms" (Meyerhoff 1999:233-234) and need to be approached and analysed as such. The localised terms of understanding gender in the Morehead area is, then, as a subset of tribal identity.

[1]It should be stressed that this finding pertains to the particular variables of vowel F1/2 and duration as investigated in this particular study. As an examiner pointed out, differences in pitch, and certain lexical differences pointed out by speakers in section 2.7 suggest that there are certain linguistic domains where gender differentiation is important. It has just happened to be that for both the vowels and [h]-drop, gender was not significant.

9.3.3 Tribe as a sociolinguistic variable

Daily bilingual women are maintaining their origin village linguistic norms, likely through sustained and continuing interaction with their origin village. The daily bilingual women are bringing Nmbo linguistic norms to their spouses and children, and their residential village speech community. Multilingualism has been raised as something that fosters variation and diversity, such as by the feature pool hypothesis (Mufwene 2001, Cheshire et al. 2011). In-marrying women continue to adhere to their tribal varieties closely, and contribute those linguistic norms into the feature pool of the speech community. The rational is, to quote Evans, "the overarching social group concurrently exposes individuals to many linguistic norm-sets, harnesses a large proportion of the semiotic space to social signalling, and gives access to a much wider range of 'linguistic memes' than would be found in a monolingual community." (2018:20).

It must be emphasised that the lack of gender as a significant variable is a finding pertaining to the Kerake tribe people of the Nmbo speech community; not to women from other tribes in the Nmbo speech community. This thesis investigated the speech of Kerake women residing in the Nen speaking village of Bimadbn, but did not systematically investigate women from other tribes marrying in to the nuclear Nmbo villages. In the qualitative descriptions I mentioned that the high salience linguistic difference between men and women is that women tend to be licensed speakers of another language (3.3.2). The fact that I the linguist, and the people of the Morehead area themselves, call these "different languages" may be obscuring the fact that another way of looking at language use in these communities is that people from different tribes and places are using a large bundle of variant linguistic makers to index their identities. If what is identified as the Nen language is treated more like a bundle of linguistic features that are slightly different to the bundle of linguistic features we call the Nmbo language, what we see in the Nmbo speech community is a high level of linguistic variation that is conditioned by tribe and place. Within a village-as-speech community view, gender and tribe will then overlap. If the studies in this thesis had included women from other tribes who speak different languages, I suspect gender and tribe would have come up as significant variables, because there is a decent chance that women will be producing linguistic features of their tribal-place variety. As mentioned in the preceding section, the localised manifestation of gender in the Morehead area is, subsumed and tied to notions of tribe. A Kerake woman is

viewed as Kerake regardless of whether she marries a man from another tribe. In this sense, exogamy and descent-group[2] as a practice-based identity (Stanford 2009) would overlap with gender, and be a significant finding of the Morehead area if we were to treat the different languages as highly codified bundles of variant linguistic forms.

To get a more holistic picture of the role of multilingualism in the Nmbo speech community, we would need to study the holistic language use of the daily bilingual women. Future work would look at the Nen speech of daily bilingual Nmbo speakers in Bimadbn, and/or to look at the Nmbo of non-Kerake women in the villages of Bevdvn and Govav. By looking at the total repertoire of these daily bilinguals, we will get a more balanced picture of multilingualism in the Morehead area, and the possible linguistic consequences of this practice. What we can say at this point in time is, however, that daily bilingual women appear quite capable of keeping up to date with the linguistic norms of their tribal language.

9.4 Individualism and Egalitarian Multilingualism, and Place-Based Variation

Earlier in the thesis I described the importance of individual autonomy in the Morehead area, and characterised the people as individualistic agents within a dense-multiplex network (2.4). I also described the Morehead area as an egalitarian multilingual language ecology (2.7). These kinds of social configurations, I believe, offer hints as to how language diversification occurs in New Guinea.

The question of how egalitarian multilingualism arises is not within the purview of this thesis, but I would like to make a few comments on it since the ethnographic details of the Morehead area fit with some observations and hypotheses made by others. Small-scale communities are often acephalous, and the relationships in small communities are thought to be egalitarian due to the lack of vertical social stratification; whether this be socio-economic stratification, or some other culturally-based power difference (e.g. spiritual power, symbolic power). Egalitarian multilingualism as an ideology is a cultural reflection of the non-stratified relationship between different groups of people. As François (2012) demonstrates, however, once the sources of social stratification make their way

[2]Although 'tribe' in the Morehead area is not specifically a system of descent, except that descent systems and lineages fall under the superstructure of the tribe.

into the lives of communities and individuals, egalitarian multilingualism as a belief can break down (also c.f. Aikhenvald 2002a). The communities of the Morehead area have traditionally been socially non-stratified (2.3). The lack of material trade suggests there was no basis of socio-economic stratification, and the use of secret words held closely within each language group suggests there were no spiritual or other cultural bases for hierarchy development. The importance placed on reciprocal sister exchange also suggests that the maintenance of egalitarian relationships across groups was (and still is) a priority. By making social exchange reciprocal, a dampener is perhaps put on the development of social stratification in the form of accumulated power in social exchange (see also Godelier 1986:xi who contrasts the Baruya of the Eastern Highlands with New Guinean communities that have reciprocal social exchange). Egalitarian multilingualism then seems to be an ideology that arises among societies that have few material or cultural basis for stratification.

Despite the connectivity afforded by social exchange, groups in the Morehead area often choose to, and are capable of, maintaining individual autonomy without reliance on other groups to fulfil material and cultural needs (2.4). Relationships with other groups are maintained for the fulfilment of the social exchange of spouses,[3] which results minimally in a loose connection of groups who are not in complete isolation from one another. The desire for autonomy and social connectivity is a constant tension for individuals and groups, but communities in the Morehead area *can* in fact maintain individual autonomy at a fundamental subsistence level without relying for support on superstructures such as the village, section, tribe, or state. The notion of ecological risk as a mechanism for explaining language diversity proposes that areas of the world with high year-round productivity have more languages because the land can support human cultural groups with smaller populations (Nettle 1999, Hua et al. 2019), and I would suggest that the peoples of the Morehead area demonstrate this theory in practice (2.2).

If we imagine shared community norms, like ideologies of tribal unity, as ties that keep groups together, these ties can be weakened for various reasons. The notion of tribe is a way of socially grouping people under some loosely recognised and shared social definition. If the social imperative is strong enough, groups

[3]Social connections are also undoubtedly maintained for affective reasons. I do not wish to down-play the role of affection and the enjoyment of others' company the Kerake express to one another, which is also undeniably part of the social landscape of the Morehead area.

can sever the shared identity that unite them, and groups can carve themselves up as distinct from others (the division between Serbia and Croatia come to mind as an example). My consultant who pondered leaving Bevdvn may start to call himself *Savarmmn är*, or 'a person of [garden place X]', should the social need ever become strong enough. Place-based linguistic variation appears to be present, and ready to go should it need to be employed for further linguistic and social differentiation.

9.5 Conclusion and Future Research

The research goal of this thesis was to investigate the sociolinguistics of language variation in a New Guinean community within an egalitarian multilingual language ecology. The research question was framed in terms of understanding the high levels of linguistic diversity we see in New Guinea today. Key findings emerged from the qualitative, linguistic, and quantitative investigations. The qualitative investigation yielded descriptions of the Nmbo speech community and language ecology. The linguistic description has produced a sketch grammar of a previously undocumented language. The quantitative findings on patterns of linguistic variability suggest the non-significance of gender, but a minor significance of village. The investigation found that the speech of Kerake Nmbo speakers pattern closely to one another, including the women who have married out of their Nmbo village of origin. The Kerake Nmbo speech community appears to be homogeneous, at least for the variables studied.

The data and findings from this thesis contribute to many subfields within linguistics. The thesis is a proof of concept that one can concurrently do both initial linguistic description and study variation. There are many lessons learnt, which I hope provide some template for future attempts at undertaking this challenging but intellectually rewarding endeavour. The sociolinguistic aspects of the thesis contribute to variationist sociolinguistics by adding insight from a "minority context"; a field that is seeing growth in recent years as the study of variation matures and is becoming a central concern in linguistics more broadly. The thesis is also a contribution to the studies of small-scale multilingualism, with presentation of primary data from speakers of an under-documented region. The sketch grammar in in line with the goals of descriptive and documentary linguistics; to gain insight into the range of linguistic structures found in the languages of the world. More specifically, the description is of a language from a notori-

ously under-described part of the world. The sketch grammar also adds another potential data point to the endeavours of comparative linguistics and typology. Finally, this thesis provides an empirical case study to the theoretical and macro-scale studies of linguistic diversity. The vowel sociophonetic and [h]-drop studies in particular pose challenges to the idea of New Guinean communities having faster tempos of change.

We have only just scratched the surface of studying variation of this intricate linguistic system, with variables from this current project mostly limited to phonetic/phonological variables. As the grammaticalisation study shows, however, the intersection of discourse with the rich morphosyntax of Nmbo is a space that promises to yield interesting findings in terms of variation and language change. The appendix also contains a list of linguistic variables which can be investigated in future studies (A.19).

There are a number of possible future studies that could build upon the findings from this thesis. One study would be to expand the investigation into the speech of the non-Kerake women who are part of the Nmbo speech community. An investigation into multilingual speech could probe which linguistic variants are used when, by whom, and how these features are categorised by their audience/interlocutor. The identification and analysis of linguistic features within such speech would yield immensely fruitful data on the usage of multiple linguistic varieties and variants within a small-scale multilingual community of New Guinea. Another possible future study is a deep investigation into the vernacular speech of a limited set of speakers while they are based at their garden hamlets. Such a study would yield important data on a more vernacular form of speech by speakers of a Southern New Guinean language.

The Nmbo speech community provides a rare case study of studying variation in a previously under-described Papuan language, with a rich and complex language ecology of egalitarian multilingualism. Through this thesis I have investigated and presented findings that mostly support prior hypotheses regarding the demographic and socio-cultural parameters responsible for fostering language diversity in New Guinea, but have added empirical data that enriches and inevitably complicates our understanding about the relationship between culture, language, and change.

Appendices

Appendix A

A.1 Sketch Grammar: Possible tautosyllabic CC combinations

Possible C_1C_2	word-initial σ	word-medial σ
/b+r/	brum 'parrot'	NA
/ ^{m}b+r/	NA	hɑ.^{m}brɑn leg/foot=LOC
/t+r/	træ.rɐh 'to scratch'	ɸɐ.trɑ.rɐh 'to break and enter'
/d+r/	dru.rɐh 'to sneeze'	?
/^{n}d+r/	NA	dro.^{n}dro 'Large-tailed nightjar bird'
/k+r/	kru.βer 'cold'	tɐ.krɑt 'hard'
/k͡p+r/	k͡prɐm.ta 'greedy'	?
/g+r/	grɑ.wɑr egret bird	wä.gruh 'to take away'
/^{ŋ}g+r/	NA	grɑ.^{ŋ}grɑ ' Little Pied Cormorant bird'
/g͡b+r/	g͡brɑ female personal name	sɑg͡brɑn tree.type=LOC
/^{ŋ}g͡b+r/	NA	?
/ɸ+r/	ɸre.rɐh 'to hold in mouth'	ɑj.nɑ.ɸro 'wild betel nut palm'
/β+r/	NA	wɑ.βroh 'to do'
/s+r/	srɑ.dɑ(male personal name)	wɑ.srɑh
/z+r/	zro^{ŋ}g yam type	?

Table A.1: Possible tautosyllabic CC combinations

A.2 Sketch Grammar: Positional Verbs

Inflected form	Variant Forms	English translation
y-ahongr		to be adorned
y-akiongr	ykiongr	to be standing
y-alongr		to be ling on ones back
y-altongr		to be under something
y-ämengr		to be hanging
y-dñingr		to be hidden
y-ehongr		to be crouched down
y-ezarungr		to be away ; be off somewhere
y-harngr		to be immersed in water
y-iätungr		to be at the end of something
y-ilawngr		to be inside
y-iyengr	iyengr	to be inclining
y-kmongr		to be in a lying position
y-marengr		to be sitting
y-ngrwn		to be approaching
y-rärngr		to be in a fork of a tree
y-rningr		to be attached to something
y-sawrngr		to be open; uncovered; unenclosed
y-tmbengr		to be close behind something
y-tongr	ytnongr	to be erected
y-uyungr		to be in a lifted up position
y-vlingr		to be inside a contained space like a container
y-vrengr		to be up high on an unstable base
y-vyongr		to be placed up high

Table A.2: List of known positional verbs in Nmbo. Inflected forms are in the 3rd singular non-prehodiernal form.

A.3 Sketch Grammar: Pronouns and Verbal Indexing Overview

	Prefixing Verb		Middle Verb
Ynd w-ngm	Ynd yn-ngm	Ynd yn-re(n)	Ynd nowav-tan
1ABS 1sg.U/go	1ABS 1nsg.U/go	1ABS 1du.U/be	1ABS 1sg.A/speak
I am going	We are going	We two are going.	I am speaking.
Bm n-ngm	Bm e-ngm	Bm e-re(n)	Bm nowav-te
2ABS 2sg.U/go	2ABS 2nsg.U/go	2ABS 2du.U/go	2ABS 2sg.A/speak
You are going	You (pl.) are going	You two are going.	You are speaking.
Bä y-ngm	Bä e-ngm	Bä e-re(n)	Bä nowav-t
1ABS 1sg.U/go	1ABS 1nsg.U/go	1ABS 1du.U/go	3ABS 3sg.A/speak
I'm going	We're going	We two are going.	S/he is speaking.

Table A.3: Absolutive pronouns and their verbal indexing.

	Transitive Verb	
yndo y\mi/tan	yndvem y\mi/tam	yndvem y\mi/wm
1sg.ERG 1sg.A>3sg.U/ask	1nsg.ERG 1nsg.A>3sg.U/ask	1nsg.ERG 1du.A>3sg.U/ask
I am asking him.	We are asking him.	We two are asking him.
ymo y\mi/t	bmovem y\mi/tate	bmovem y\mi/te
2sg.ERG 2sg.A>3sg.U/ask	2nsg.ERG 2nsg.A>3sg.U/ask	2nsg.ERG 2du.A>3sg.U/ask
You are asking him.	You (pl.) are asking him.	You two are asking him.
ymo y\mi/t	ymovem y\mi/tat	ymovem y\mi/wt
3sg.ERG 3sg.A>3sg.U/ask	3nsg.ERG 3nsg.A>3sg.U/ask	3nsg.ERG 3du.A>3sg.U/ask
S/he is asking him.	They are asking him.	Those two are asking him.

Table A.4: Ergative pronouns, and their verbal indexing.

A.4 NWC: Full Nmbo Word List Words By Year of Collection

	2014	2015	2016 + 2017
i	kaki	qɐmbiv, qɐki kaki, bitræri, tikv, wingoh	fivi, qɐki, itroh tikv ziwæl qɐmbiv
e	deve, hure bermber, tande	deve, sonzær bwe, hure tekr, hrare, mer, qɐvte bwe, fader, mer, evh	deve, end
æ	bæ, dæv bærmbær	ægævu, sonzær bwe, sgær fætr-fætr bæ	budær, wæz svælnḡ, ziwæl
a	anu, kaki, tande	hakr, kaki, kaka adodo, hrare hamba, fader, sawa	zav, ḡarnḡar, kaka barmbar, hhnar, orang odaga, hamba, band, fan, adodo
o	bombo	got, sonzær bwe, adodo, tot weikor, tok, mngo, wingoh	itroh, odaga, orang adodo, tok
u	hure, anu	ægævu, butɐm hure, budær	budær, zuzu butəm, hvu
ə		butəm sgær, ḡəḡən	butɐm, bərbər ḡəḡən
ɐ	bm knm, hkv,	bɐm kɐnɐm, hɐkɐv, hɐrare, tikɐv, mɐngo,evɐh, qt, kɐkɐv, hɐkr, qɐmbiv ,qɐki	ynḡhə, kəv, svælnḡ tikɐv, hɐhɐnar, bɐm kɐnɐm, hɐvu bɐrbɐr, qɐki, qɐmbiv, sɐq

A.5 Additional Comments for Nmbo Grammar

On the manner demonstratives (section 4.2.1.3)

The manner demonstratives *dena* and *mna* likely arose from temporal particles. The particles *de* and *mé* in contemporary use function as a completive (section 4.3.1.2) and continuative (section 4.3.1.4) respectively, and one can see how the semantics of this overlap with discourse functions of 'look back' and 'anticipation'. One is tempted to hypothesise that the pronominal *yna*, also ending in a segment *-na* arose in part from the dative 3rd singular *yä*. This may explain why the distal *kt* exhibits irregular behaviour in contemporary case marking paradigms, where it is unable to take certain case markers such as the goal suffix, while its semantic counterpart *yna* can. In the sister language of Nen the corresponding words are *dne*, *mne*, and *yne*, suggesting the source form is *ne* and/or *na*.

Demonstrative	Composition	Relevant free word
dena	de-na	*de* (completive particle)
mna	m-na	*mé* (continuative particle)
yna	y-na	*yä* (3rd singular dative)
kt	NA	NA

Table A.5: Relationship between demonstratives. NA = not available

A.6 NSC: Speakers, speech amount, and token count

Name	Code	YoB	Sex	VoO	VoR	Speech Type Type	Duration (mins)	[h]-word	DEM+COP
Mesly	yfggMW	1997	F	Govav	Govav	Coconut	03:45	3	2
						MQ	03:30	2	2
						Cassoway	05:45	21	5
						Other	10:25	8	9
Total							23:25	34	18
Gabrielle	ymggGW	1994	M	Govav	Govav	MQ	04:20	5	5
						Cassoway	07:30	48	9

Table A.6 – continued from previous page

Name	Code	YoB	Sex	VoO	VoR	Speech Type	Duration (mins)	[h]-word	DEM+COP
						Other	04:15	8	4
Total							16m 5s	61	18
Manika	ymbbME	1997	M	Bevdvn	Bevdvn	MQ	01:10	0	0
						Cassoway	03:40	13	0
						Other	01:08	0	0
Total							5m 58s	13	0
Bvk	ymggBS	1994	M	Govav	Govav	MQ	05:15	6	2
						Cassowary	08:00	60	7
						Other	10:30	2	9
						Total	13m 15s	66	9
Ambo	yfbbAY	1990	F	Bevdvn	Bevdvn	MQ	00:58	0	0
						Cassowary	02:50	34	1
						Other	02:13	4	0
Total						Total	3m 48s	34	1
Bethsheba	yfgmBZ	1989	F	Govav	Bimadbn	Coconut	01:15	5	3
						MQ	02:00	1	0
						Cassowary	06:20	46	4
						Other	04:15	1	1
Total							9m 35s	52	7
Rachel	yfabRM	1989	F	Bevdvn	Bevdvn	MQ	01:30	3	3
						Cassowary	07:30	20	3
						Other	19:17	23	15
Total							28m 17s	46	21
Smako Jr	ymmbSL	1989	M	Bevdvn	Bevdvn	MQ	05:00	1	1
						Cassowary	06:50	25	2
						Other	06:41	3	2
Total							18m 31s	29	5
Jacklyn	yfbbJY	1986	F	Bevdvn	Bevdvn	Coconut	02:07	3	3
						MQ	00:58	1	1
						Cassowary	06:26	22	5
						Other	15:20	19	10
Total							9m 31s	26	9
Alice	yfbbAT	1989	F	Bevdvn	Bevdvn	Other	20:20	34	12

Table A.6 – continued from previous page

Name	Code	YoB	Sex	VoO	VoR	Speech Type	Duration (mins)	[h]-word	DEM+COP
Total							20m 20s	34	12
Mada	ymbbMZ	1985	M	Bevdvn	Bevdvn	Other	28m 43s	16	12
Total							28m 43s	16	12
Richard	mmggRG	1985	M	Govav	Govav	Coconut	03:20	5	4
						MQ	13:25	10	6
						Cassowary	14:20	42	5
						Other	10:41	6	3
Total							31m 5s	57	15
Kata	mmbbKS	1979	M	Bevdvn	Bevdvn	Coconut	05:15	0	0
						MQ	02:05	3	4
						Cassowary	05:38	24	11
						Other	19:20	24	22
Total							32:18	51	37
GimaZ	mmggGZ	1984	M	Govav	Govav	MQ	04:20	5	7
						Cassowary	06:08	12	7
						Other	19:16	16	9
Total							29m 44s	33	23
Thomas	ymbbTT	1982	M	Bevdvn	Bevdvn	MQ	01:40	6	5
						Cassowary	04:55	20	4
						Other	20:20	17	16
Total							26m 55s	43	25
Kawas	mmbbKS2	1977	M	Bevdvn	Bevdvn	Coconut	04:20	5	5
						Cassowary	04:26	17	7
						Other	26:37	19	14
Total							35m 23s	41	26
Mära	mfggMW	1968	M	Govav	Govav	MQ	07:00	7	17
						Cassowary	10:20	39	8
						Other	04:30	1	6
Total							7m 0s	7	17
Alqi	mfgbAN	1970	F	Govav	Bevdvn	MQ	04:05	9	4
						Cassowary	04:07	40	0
						Other	08:55	11	0
Total							17m 7s	60	4

Table A.6 – continued from previous page

Name	Code	YoB	Sex	VoO	VoR	Speech Type	Duration (mins)	[h]-word	DEM+COP
Rhouda	mfaaRT	1969	F	Bevdvn	Bevdvn	MQ	01:54	3	3
						Cassowary	03:19	15	9
						Other	10m 41s	28	11
Total						Total	15m 54s	46	23
Gvai	mmggGS	1976	M	Govav	Govav	Coconut	02:00	1	0
						Other	22:20	12	4
Total							24m 20s	13	4
Gwnyo	mfgmGwN	1965	F	Govav	Bimadbn	Coconut	04:10	2	6
						MQ	11:04	7	8
						Cassoway	04:15	30	2
						Other	06:56	7	15
Total							26m 25s	46	31
Yayam	mfbmYS	1972	F	Bevdvn	Bimadbn	Coconut	05:10	7	11
						MQ	05:20	4	1
						Cassowary	04:16	25	1
						Other	16:58	22	34
Total							31m 44s	58	47
Bunai	mmaaBA	1969	M	Bevdvn	Bevdvn	Other	33m 49s	57	14
Total							33m 49s	57	14
Lazarus	mmbbLS	1967	M	Bevdvn	Bevdvn	Coconut	07:40	9	4
						MQ	02:40	4	1
						Other	14:43	13	6
Total							25m 3s	26	11
Mwigm	mmggMN	1968	M	Govav	Govav	Coconut	04:29	1	3
						MQ	10:10	6	3
						Cassowary	06:00	38	4
						Other	23:00	24	10
Total							43m 39s	45	10
ZogaG	smggZG	1959	M	Govav	Govav	Coconut	07:00	8	0
						MQ	09:20	0	0
						Cassowary	10:25	30	9
						Other	13:55	14	3
Total							40m 40s	52	12

Table A.6 – continued from previous page

Name	Code	YoB	Sex	VoO	VoR	Speech Type	Duration (mins)	[h]-word	DEM+COP
Gerida	mfgmGN	1966	F	Govav	Bimadbn	Coconut	07:30	5	24
						MQ	07:10	8	4
						Cassowary	07:15	43	21
						Other	07:20	1	4
Total							29m 15s	57	53
Wäwi	sfbaWZ	1957	F	Bevdvn	Bevdvn	Coconut	02:30	0	5
						MQ	05:00	7	3
						Other	13:20	22	2
Total							20m 50s	29	10
Ruscien	sfamRA	1960	F	Bevdvn	Bimadbn	Coconuts	14:02	15	25
						MQ	10:20	11	4
						Cassowary	05:00	19	6
						Other	08:50	18	7
Total							38m 12s	63	42
Fangore	sfgmFY	1955	F	Govav	Bimadbn	Coconut	05:40	3	4
						MQ	10:40	7	11
						Cassowary	08:25	36	7
						Other	11:14	13	8
Total							35m 59s	59	30
GimaM	smggGM	1954	M	Govav	Govav	Coconut	03:30	1	5
						Other	24:36	28	30
Total							28m 6s	29	35
Taqm	smbbTS	1953	M	Bevdvn	Bevdvn	MQ	06:20	9	0
						Cassowary	08:40	22	2
						Other	31:33	17	16
Total							46m 33s	48	18
Dorissa	sfggDS	1952	F	Govav	Govav	MQ	05:15	8	1
						Cassowary	06:45	28	1
						Other	11:20	15	4
Total							23m 20s	51	6
Yavs	smbbYZ	1951	M	Bevdvn	Bevdvn	Coconut	02:00	2	4
						MQ	06:40	8	7
						Cassowary	05:00	55	6

Table A.6 – continued from previous page

Name	Code	YoB	Sex	VoO	VoR	Speech Type	Duration (mins)	[h]-word	DEM+ COP
						Other	17:03	10	10
Total							30m 43s	75	27
Masa	smggMG	1954	M	Govav	Govav	Coconut	02:40	2	2
						MQ	15:20	3	2
						Cassowary	09:05	32	8
						Other	12:17	1	1
Total							39m 22s	38	13
Maiwa	smggMY	1952	M	Govav	Govav	MQ	07:40	5	12
						Cassowary	07:10	33	1
						Other	06:47	4	2
Total							21m 37s	42	15
Hom	sfbfHZ	1949	F	Bevdvn	Bimadbn	Other	08:55	14	13
Total							8m 55s	14	13
Smako	smbbSZ	1948	M	Bevdvn	Bevdvn	Coconut	02:28	1	0
						MQ	07:40	0	0
						Cassowary	05:40	18	7
						Other	22:23	31	9
Total							43m 51s	68	23
Total							15:56:22	1605	683

A.7 NSC: Content and sample questions

Coconut Interview
Who planted this coconut?
Why did you/they plant this coconut?
What do you do with the coconut leaves?
What do you do with the coconut fronds?
What do you do with the coconuts themselves?
What will happen to this coconut tree when you pass away? (i.e. who inherits it for use)
Multilingual Questionnaire
Where were you born?
What language does your mother speak?
Did you speak your mother's language with her?
When did you speak your mother's langauge?
What language do you speak to your mother's brothers?
Did you visit your mother's village as a child?
What language do you speak when you visit Bimadebn/Keru/Mata etc.
What langauges can you speak?
What languages can you hear?
What languages are you no good at?
Biographical Questions
Did you attend school?
Have you travelled around or have you only lived in your village?
What games did you play as a child?
Who did you play with as a child?
What activities/chores did you do as a child?
Perceptual questions
What languages are spoken in X location?
How do you feel when you hear non-Kerake speaking Nmbo?
How do you feel when you hear Kerake people speaking languages other than Nmbo?
Do you think men and women speak differently?

How do you think young people speak Nmbo?

Why are there so many languages in the Morehead area?

Marriage Stories

Near-Death Stories

Ethnographic Descriptions

Cassowary Picture Task

Explain what you see.

Retell the story from a third person perspective to children who are physically present.

A.8 S1: Formulae

The Lobanov transformation formula applied via the Vowel package is from Adank, Smits, and van Hout (2004):

$$F_{ti} = \frac{F_{ti} - \mu_{ti}}{\delta_{ti}}$$

μ_{ti} refers to the average formant frequency across all vowels for talker t, δ_{ti} refers to the standard deviation of average μ_{ti} (Adank, Smits, and van Hout 2004:3101).

For duration calculations, speech rate was normalised using the formula presented in Wassink (2006) (p.2345). The mean duration of each vowel category ([i, e, æ, ɑ, o u, ə, ɐ]) is calculated for each speaker. A grand mean of duration for each speaker is then calculated across all vowel categories. This is done by the following formula, where $\overline{D}_{o,k}$ is the grand mean of duration for speaker k, and D_{ijk} is the observed segment duration for token i of vowel j for speaker k. n_j is the number of observations per vowel j.

$$\overline{D}_{o,k} = \frac{\sum_{i=1}^{n}(\sum_{i=1}^{n} D_{ijk}/n_j)}{n}$$

This grand mean is then subtracted from each observed segment duration.

$$\delta_{ijk} = D_{ijk} - \overline{D}_{ok}$$

Means per speaker for F1, F2, and duration are all calculated by combining measurement of vowels occurring in all environments, i.e. word initial, final, intervocalic, in stressed and unstressed positions. The phonetic environments of the words are then tested as fixed effects in the linear mixed models, in order to determine whether they have an effect on the vowel parameters.

Vowel space spread is determined by calculating the area of a trapezium formed by the extremity vowels per speaker. The parameters are the mean F2 values for [i, u, æ, a] for the bases, and mean F1 values for [i, a] for height. In other words a = uF2 - iF2, b = aF2 - æF2, h = aF1 - iF1.

$$\frac{a+b}{2} \times h$$

A.9 S1: Obstruents Only Word List

Vowel	Words
i	b**i**.træ.r**i**, ka.k**i**, t**i**.kv, f**i**.v**i**, z**i**.wæl
e	d**e**.v**e**, d.ma.v**e**
æ	æ.g**æ**.vu, b**æ**, bi.træ.ri, bu.d**æ**r, f**æ**tr.f**æ**tr, s.g**æ**r, son.z**æ**r.bwe, s.v**æ**.lng̅
u	æ.gæ.v**u**, b**u**.dær, b**u**.təm
o	a.d**o**.d**o**, g**o**t, mn.g**o**, s**o**n.zær.bwe, t**o**k, t**o**t, wei.k**o**r, b**o**m.b**o**, **o**.rng
a	k**a**.k**a**, k**a**.ki, k**a**.n**a**m, s**a**.wa, sr.v**a**v, z**a**v, f**a**.der
ə	bu.t**é**m, g̅**é**.g̅**é**n, s**é**.gær, b**é**r.b**é**r, b**é**.rm.b**é**r
ɐ	b**ɐ**m.k**ɐ**.ɐm, e.v**ɐ**h, , k**ɐ**.k**ɐ**v, q**ɐ**.ki q**ɐ**m.biv, q**ɐ**t, q**ɐ**v, ti.k**ɐ**v, s**ɐ**q , s**ɐ**r.vav, s**ɐ**.væ.lɐng̅, t**ɐ**nd, b**ɐ**.rm.b**ɐ**r

Table A.8: Subset of words from NWC where target vowel follows obstruents.

A.10 S1: Vowel token count by Speaker

Table A.9: Vowel token count by Speaker

Speaker	i	e	æ	u	o	a	ə	ɐ	Sex	Yob	VoO	VoR
sfggDS	21	12	27	9	27	24	12	36	F	1947	Govav	Govav
sfgbMM	28	15	29	18	27	24	33	58	F	1949	Govav	Bevdvn
sfggZS	35	18	36	17	27	21	27	53	F	1949	Govav	Bevdvn
sfbgKS	21	12	27	8	27	24	12	33	F	1953	Bevdvn	Govav
sfgmFY	21	13	21	12	14	16	18	36	F	1955	Govav	Bimadebn
sfbbKS	36	24	44	14	39	39	24	68	F	1956	Bevdvn	Bevdvn
sfabKK	21	15	27	9	26	26	11	34	F	1957	Bevdvn	Bevdvn
sfabWP	18	12	27	9	27	24	12	33	F	1957	Bevdvn	Bevdvn
sfbaWZ	51	18	41	25	45	29	45	71	F	1957	Bevdvn	Bevdvn
sfbbDS	24	18	33	12	33	27	24	48	F	1957	Bevdvn	Bevdvn
sfamRA	21	12	22	12	18	18	21	42	F	1960	Bevdvn	Bimadebn
mfgmGwN	21	12	22	12	16	18	20	41	F	1965	Govav	Bimadebn
sfgdWS	19	12	23	8	13	21	8	27	F	1968	Govav	Govav
mfggMW	21	12	21	12	18	18	21	33	F	1969	Govav	Govav
mfagWA	21	6	27	9	27	24	12	35	F	1970	Bevdvn	Govav
mfgbAN	42	24	49	20	43	40	33	77	F	1970	Govav	Bevdvn
mfbmYS	21	12	21	12	16	18	21	40	F	1972	Bevdvn	Bimadebn
mfbmHY	21	12	21	12	17	18	21	33	F	1976	Bevdvn	Bimadebn
yfggJW	21	12	27	8	27	24	12	36	F	1985	Govav	Govav
yfbbJY	48	30	66	23	50	39	47	83	F	1986	Bevdvn	Bevdvn
yfggED	21	12	27	9	22	18	12	33	F	1987	Govav	Govav
yfabRM	42	23	48	21	45	41	27	75	F	1989	Bevdvn	Bevdvn
yfbbAT	24	18	33	12	28	27	24	45	F	1989	Bevdvn	Bevdvn
yfggGY	18	6	15	6	15	12	6	18	F	1989	Govav	Govav
yfgmBZ	21	12	21	12	17	17	21	42	F	1990	Govav	Bimadebn
yfbbYL	38	24	52	24	51	39	45	75	F	1991	Bevdvn	Bevdvn
yfggMW	21	12	27	9	26	24	12	36	F	1992	Govav	Govav
smggWK	21	12	18	12	21	18	21	33	M	1943	Govav	Govav
smmgWW	21	12	21	12	18	18	21	30	M	1943	Govav	Govav

Table A.9 – continued from previous page

Speaker	i	e	æ	u	o	a	ə	ɐ	Sex	Yob	VoO	VoR
smggWG	39	22	48	20	44	41	32	72	M	1947	Govav	Govav
smbbSZ	48	32	59	24	52	44	39	91	M	1948	Bevdvn	Bevdvn
smggZG	42	15	41	18	39	30	33	64	M	1949	Govav	Govav
smbbYZ	47	36	51	29	47	42	45	103	M	1951	Bevdvn	Bevdvn
smbbTS	39	24	41	15	45	42	30	72	M	1953	Bevdvn	Bevdvn
smggGM	24	15	33	11	27	27	24	42	M	1954	Govav	Govav
smggMG	48	24	53	15	48	39	24	75	M	1954	Govav	Govav
mmbbJS	39	21	38	15	42	42	24	69	M	1962	Bevdvn	Bevdvn
mmbbLS	27	15	34	18	24	27	21	53	M	1967	Bevdvn	Bevdvn
mmggMN	45	24	47	20	51	45	44	81	M	1968	Govav	Govav
mmbbES	19	9	27	9	26	27	12	33	M	1970	Bevdvn	Bevdvn
mmbbST	21	15	27	9	27	27	12	36	M	1971	Bevdvn	Bevdvn
mmbbTY	21	14	27	9	27	26	11	35	M	1973	Bevdvn	Bevdvn
mmggWW	21	12	21	12	21	18	27	33	M	1973	Govav	Govav
mmbbBD	26	21	27	9	22	27	21	51	M	1975	Bevdvn	Bevdvn
mmbbLM	44	27	42	18	45	42	36	86	M	1975	Bevdvn	Bevdvn
mmggGN	21	12	21	9	21	18	18	39	M	1975	Govav	Govav
mmggGS	21	12	21	12	21	18	21	39	M	1976	Govav	Govav
mmbbKS2	20	12	26	9	27	22	10	36	M	1977	Bevdvn	Bevdvn
mmbbKS	39	21	54	24	44	30	33	78	M	1979	Bevdvn	Bevdvn
mmbbKS3	21	12	27	9	27	26	12	33	M	1979	Bevdvn	Bevdvn
mmbbSS	21	15	27	9	26	27	12	36	M	1980	Bevdvn	Bevdvn
mmggDZ	21	12	21	12	21	18	21	36	M	1980	Govav	Govav
ymbbTT	36	20	45	21	45	36	30	72	M	1982	Bevdvn	Bevdvn
ymbbMY	36	15	42	20	38	33	32	68	M	1983	Bevdvn	Bevdvn
mmggGZ	42	27	47	21	51	45	33	85	M	1984	Govav	Govav
mmggRG	42	24	36	15	42	42	33	80	M	1985	Govav	Govav
ymbbMZ	20	9	27	9	24	24	12	36	M	1985	Bevdvn	Bevdvn
ymggSY1	21	9	27	9	27	24	12	36	M	1985	Govav	Govav
mmggRW	21	12	21	12	21	18	21	37	M	1989	Govav	Govav

Table A.9 – continued from previous page

Speaker	i	e	æ	u	o	a	ə	ɐ	Sex	Yob	VoO	VoR
ymmbSL	21	12	21	12	21	18	21	39	M	1989	Bevdvn	Bevdvn
ymggSY2	21	9	27	9	27	24	12	36	M	1991	Govav	Govav
ymggBS	28	15	26	15	28	23	30	56	M	1994	Govav	Govav
ymggGW	50	24	49	22	51	47	41	86	M	1994	Govav	Govav

A.11 S1: Individual Speaker F1 and F2 Means

Table A.10: Speaker F1 means per vowel (Lobanov normalised)

Speaker	Sex	OriV	Yob	i	e	æ	u	o	a	ə	ɐ
mfagWA	F	Bevdvn	1970	0.206	1.035	1.293	0.143	0.711	1.484	0.848	1.139
mfbmHY	F	Bevdvn	1976	0.226	0.670	1.334	0.253	0.750	1.888	0.838	1.088
mfbmYS	F	Bevdvn	1972	0.280	0.661	1.271	0.353	0.881	1.580	0.925	1.216
mfgbAN	F	Govav	1970	0.245	0.735	1.359	0.284	0.741	1.592	0.839	1.110
mfggMW	F	Govav	1969	0.104	0.849	1.490	0.170	0.646	1.541	1.058	1.166
mfgmGwN	F	Govav	1965	0.087	0.593	1.508	0.095	0.477	1.673	0.934	1.211
mmbbBD	M	Bevdvn	1975	0.090	0.724	1.365	0.115	0.660	1.680	0.773	1.089
mmbbES	M	Bevdvn	1970	0.288	0.731	1.314	0.184	0.681	1.663	0.909	0.847
mmbbJS	M	Bevdvn	1962	0.228	0.725	1.506	0.254	0.719	1.596	0.780	0.974
mmbbKS	M	Bevdvn	1979	0.271	0.880	1.447	0.350	0.793	1.598	0.903	0.979
mmbbKS2	M	Bevdvn	1977	0.403	0.876	1.439	0.400	0.562	1.502	0.786	0.985
mmbbKS3	M	Bevdvn	1979	0.251	0.851	1.427	0.318	0.756	1.476	0.849	0.925
mmbbLM	M	Bevdvn	1975	0.209	0.954	1.350	0.403	0.830	1.548	0.861	1.071
mmbbLS	M	Bevdvn	1967	0.232	0.779	1.372	0.314	0.805	1.638	0.824	1.073
mmbbSS	M	Bevdvn	1980	0.468	0.956	1.345	0.428	0.838	1.403	0.623	0.899
mmbbST	M	Bevdvn	1971	0.265	0.816	1.299	0.271	0.783	1.650	0.644	0.968
mmbbTY	M	Bevdvn	1973	0.148	0.957	1.329	0.171	0.713	1.682	0.663	0.889
mmggDZ	M	Govav	1980	0.099	0.908	1.407	0.163	0.775	1.427	0.976	1.224
mmggGN	M	Govav	1975	0.011	0.789	1.430	0.037	0.838	1.483	0.921	1.266
mmggGS	M	Govav	1976	0.189	0.904	1.395	0.331	0.880	1.567	0.898	1.103
mmggGZ	M	Govav	1984	0.104	0.700	1.385	0.274	0.785	1.625	0.852	1.085
mmggMN	M	Govav	1968	0.120	0.808	1.432	0.224	0.799	1.561	0.913	1.059
mmggRG	M	Govav	1985	0.101	0.790	1.462	0.177	0.785	1.601	0.877	1.093
mmggRW	M	Govav	1989	0.042	0.825	1.545	0.151	0.741	1.653	0.821	1.192
mmggWW	M	Govav	1973	0.111	0.814	1.485	0.172	0.537	1.861	0.743	1.136
sfabKK	F	Bevdvn	1957	0.074	0.783	1.241	0.114	0.680	1.630	0.768	1.163
sfabWP	F	Bevdvn	1957	0.155	0.740	1.223	0.160	0.691	1.606	0.546	1.171
sfamRA	F	Bevdvn	1960	0.175	0.796	1.358	0.256	0.800	1.736	0.759	1.160
sfbaWZ	F	Bevdvn	1957	0.155	0.607	1.500	0.237	0.782	1.797	0.863	1.265

Table A.10 – continued from previous page

Speaker	Sex	OriV	Yob	i	e	æ	u	o	a	ə	ɐ
sfbbDS	F	Bevdvn	1957	0.102	0.874	1.333	0.111	0.552	1.648	0.922	1.095
sfbbKS	F	Bevdvn	1956	0.090	0.682	1.422	-0.006	0.687	1.610	0.872	1.101
sfbgKS	F	Bevdvn	1953	0.290	0.661	1.228	0.363	0.713	1.767	0.783	0.913
sfgbMM	F	Govav	1949	0.239	0.745	1.453	0.263	0.868	1.626	0.826	1.080
sfgdWS	F	Govav	1968	0.128	0.877	1.299	0.131	0.669	1.646	0.708	0.971
sfggDS	F	Govav	1947	0.194	0.839	1.441	0.185	0.720	1.491	0.767	1.027
sfggZS	F	Govav	1949	0.414	0.871	1.425	0.411	0.792	1.548	0.968	1.017
sfgmFY	F	Govav	1955	0.186	0.655	1.552	0.260	0.569	1.513	0.976	1.153
smbbSZ	M	Bevdvn	1948	0.336	0.765	1.568	0.263	0.565	1.759	0.840	0.834
smbbTS	M	Bevdvn	1953	0.217	0.672	1.454	0.191	0.618	1.710	0.851	1.023
smbbYZ	M	Bevdvn	1951	0.266	0.783	1.491	0.231	0.672	1.706	0.717	1.070
smggGM	M	Govav	1954	0.162	0.834	1.375	0.177	0.704	1.596	0.702	1.078
smggMG	M	Govav	1954	0.193	0.730	1.478	0.097	0.735	1.366	1.017	1.152
smggWG	M	Govav	1947	0.074	0.700	1.450	0.004	0.450	1.581	0.921	1.217
smggWK	M	Govav	1943	0.173	0.685	1.537	0.177	0.616	1.490	1.026	1.285
smggZG	M	Govav	1949	0.113	0.559	1.443	0.208	0.650	1.713	0.903	1.240
smmgWW	M	Govav	1943	0.378	0.847	1.550	0.415	0.588	1.617	1.045	0.847
yfabRM	F	Bevdvn	1989	0.181	0.884	1.352	0.141	0.754	1.651	0.807	1.060
yfbbAT	F	Bevdvn	1989	0.180	0.892	1.269	0.133	0.765	1.539	0.807	1.170
yfbbJY	F	Bevdvn	1986	0.227	0.812	1.356	0.298	0.776	1.614	0.912	1.078
yfbbYL	F	Bevdvn	1991	0.224	0.881	1.399	0.301	0.699	1.588	0.876	1.081
yfggED	F	Govav	1987	0.235	0.984	1.426	0.251	0.738	1.277	0.830	1.047
yfggGY	F	Govav	1989	0.134	0.767	1.522	0.258	0.740	1.860	1.060	0.863
yfggJW	F	Govav	1985	0.117	0.775	1.347	0.311	0.798	1.374	0.870	1.153
yfggMW	F	Govav	1992	0.125	0.714	1.365	0.184	0.777	1.594	0.746	1.059
yfgmBZ	F	Govav	1990	0.247	0.609	1.338	0.480	0.709	1.658	0.989	1.107
ymbbMY	M	Bevdvn	1983	0.305	0.898	1.180	0.345	1.088	1.541	0.802	1.065
ymbbMZ	M	Bevdvn	1985	0.301	0.886	1.404	0.380	0.751	1.604	0.607	0.894
ymbbTT	M	Bevdvn	1982	0.256	0.892	1.352	0.322	0.864	1.528	0.914	1.025
ymggBS	M	Govav	1994	0.246	0.863	1.641	0.309	0.765	1.588	0.872	0.929

Table A.10 – continued from previous page

Speaker	Sex	OriV	Yob	i	e	æ	u	o	a	ə	ɐ
ymggGW	M	Govav	1994	0.218	0.884	1.455	0.190	0.681	1.651	0.895	1.007
ymggSY1	M	Govav	1985	0.184	0.904	1.401	0.304	0.858	1.427	0.797	1.005
ymggSY2	M	Govav	1991	0.237	0.904	1.347	0.378	0.788	1.477	0.941	1.016
ymmbSL	M	Bevdvn	1989	0.223	0.719	1.520	0.318	0.730	1.709	0.852	1.068

Table A.11: Speaker F2 means per vowel (Lobanov normalised)

Speaker	Sex	OriV	Yob	i	e	æ	u	o	a	ə	ɐ
mfagWA	F	Bevdvn	1970	1.861	1.637	1.193	0.133	0.351	0.813	1.082	0.693
mfbmHY	F	Bevdvn	1976	1.813	1.413	1.036	0.138	0.302	0.744	1.036	0.705
mfbmYS	F	Bevdvn	1972	1.817	1.508	1.144	0.185	0.212	0.703	1.019	0.684
mfgbAN	F	Govav	1970	1.787	1.636	1.205	0.103	0.111	0.722	1.122	0.673
mfggMW	F	Govav	1969	1.828	1.650	1.099	0.027	0.233	0.572	1.072	0.700
mfgmGwN	F	Govav	1965	1.711	1.604	1.193	0.047	0.038	0.630	1.255	0.653
mmbbBD	M	Bevdvn	1975	1.849	1.568	1.087	0.296	0.126	0.726	1.064	0.580
mmbbES	M	Bevdvn	1970	1.904	1.628	1.261	0.188	0.269	0.716	1.298	0.562
mmbbJS	M	Bevdvn	1962	1.715	1.628	1.211	0.335	0.336	0.748	1.101	0.644
mmbbKS	M	Bevdvn	1979	1.906	1.519	1.147	0.340	0.232	0.583	1.129	0.670
mmbbKS2	M	Bevdvn	1977	1.875	1.552	1.209	0.061	0.333	0.821	1.137	0.591
mmbbKS3	M	Bevdvn	1979	2.174	1.703	1.233	0.264	0.274	0.418	0.964	0.418
mmbbLM	M	Bevdvn	1975	1.872	1.536	1.208	0.256	0.233	0.662	1.144	0.587
mmbbLS	M	Bevdvn	1967	1.876	1.551	1.212	0.154	0.234	0.666	1.112	0.667
mmbbSS	M	Bevdvn	1980	1.883	1.701	1.271	0.264	0.191	0.604	1.037	0.524
mmbbST	M	Bevdvn	1971	1.825	1.467	1.247	0.252	0.349	0.619	1.265	0.631
mmbbTY	M	Bevdvn	1973	1.734	1.502	1.268	0.528	0.267	0.807	1.086	0.593
mmggDZ	M	Govav	1980	2.075	1.753	1.071	0.124	0.157	0.546	0.994	0.577
mmggGN	M	Govav	1975	1.876	1.657	1.061	0.139	0.195	0.700	0.977	0.702
mmggGS	M	Govav	1976	2.018	1.531	0.953	0.231	0.338	0.697	0.939	0.711
mmggGZ	M	Govav	1984	1.862	1.569	1.211	0.279	0.264	0.692	1.137	0.677

Table A.11 – continued from previous page

Speaker	Sex	OriV	Yob	i	e	æ	u	o	a	ə	ɐ
mmggMN	M	Govav	1968	1.814	1.479	1.124	0.276	0.287	0.753	1.150	0.726
mmggRG	M	Govav	1985	1.945	1.689	1.110	0.072	0.174	0.632	1.083	0.617
mmggRW	M	Govav	1989	1.915	1.361	1.076	0.347	0.260	0.725	0.969	0.721
mmggWW	M	Govav	1973	1.928	1.533	1.032	0.260	0.254	0.682	1.042	0.711
sfabKK	F	Bevdvn	1957	1.884	1.473	1.071	0.484	0.303	0.752	1.040	0.645
sfabWP	F	Bevdvn	1957	1.920	1.609	1.070	0.565	0.229	0.715	1.228	0.704
sfamRA	F	Bevdvn	1960	1.857	1.488	1.183	0.318	0.257	0.638	1.101	0.591
sfbaWZ	F	Bevdvn	1957	1.893	1.703	1.070	0.178	0.038	0.612	1.043	0.626
sfbbDS	F	Bevdvn	1957	1.912	1.569	1.215	0.025	0.111	0.721	1.207	0.660
sfbbKS	F	Bevdvn	1956	1.656	1.579	1.171	0.298	0.189	0.795	1.099	0.699
sfbgKS	F	Bevdvn	1953	2.015	1.710	1.161	0.393	0.246	0.617	1.110	0.459
sfgbMM	F	Govav	1949	1.835	1.704	1.164	0.295	0.177	0.723	1.028	0.694
sfgdWS	F	Govav	1968	1.655	1.490	1.249	0.285	0.114	0.640	1.245	0.556
sfggDS	F	Govav	1947	1.896	1.616	1.151	0.444	0.256	0.635	1.180	0.653
sfggZS	F	Govav	1949	1.929	1.823	1.020	0.210	0.128	0.442	0.967	0.459
sfgmFY	F	Govav	1955	1.879	1.635	0.952	0.076	0.018	0.712	0.986	0.697
smbbSZ	M	Bevdvn	1948	1.603	1.583	1.236	0.450	0.223	0.774	1.082	0.693
smbbTS	M	Bevdvn	1953	1.864	1.686	1.165	0.111	0.128	0.691	1.068	0.673
smbbYZ	M	Bevdvn	1951	1.699	1.537	1.244	0.339	0.207	0.680	1.145	0.690
smggGM	M	Govav	1954	1.869	1.596	1.144	0.388	0.188	0.605	1.173	0.629
smggMG	M	Govav	1954	1.848	1.581	1.135	0.180	0.165	0.589	1.196	0.619
smggWG	M	Govav	1947	1.678	1.542	1.138	0.363	0.319	0.740	1.106	0.776
smggWK	M	Govav	1943	1.662	1.483	1.139	0.235	0.145	0.719	1.270	0.800
smggZG	M	Govav	1949	1.734	1.414	1.178	0.112	0.225	0.700	1.236	0.692
smmgWW	M	Govav	1943	1.722	1.634	1.203	0.508	0.367	0.571	1.043	0.439
yfabRM	F	Bevdvn	1989	1.966	1.565	1.154	0.247	0.254	0.661	0.991	0.634
yfbbAT	F	Bevdvn	1989	1.801	1.589	1.058	0.224	0.296	0.794	1.023	0.727
yfbbJY	F	Bevdvn	1986	1.999	1.716	1.096	0.126	0.158	0.595	0.993	0.558
yfbbYL	F	Bevdvn	1991	1.891	1.683	1.226	0.042	0.214	0.692	1.093	0.659
yfggED	F	Govav	1987	1.845	1.601	1.156	0.332	0.340	0.599	0.957	0.558

Table A.11 – continued from previous page

Speaker	Sex	OriV	Yob	i	e	æ	u	o	a	ə	ɐ
yfggGY	F	Govav	1989	1.886	1.462	1.121	0.206	0.198	0.604	1.154	0.431
yfggJW	F	Govav	1985	1.961	1.604	1.183	0.144	0.221	0.597	1.393	0.542
yfggMW	F	Govav	1992	1.999	1.599	1.145	0.523	0.319	0.600	1.077	0.568
yfgmBZ	F	Govav	1990	1.826	1.530	1.089	0.181	0.265	0.671	1.083	0.644
ymbbMY	M	Bevdvn	1983	1.805	1.531	1.120	0.426	0.444	0.748	1.060	0.683
ymbbMZ	M	Bevdvn	1985	1.928	1.618	1.161	0.229	0.412	0.766	0.975	0.597
ymbbTT	M	Bevdvn	1982	1.881	1.519	1.144	0.282	0.357	0.748	1.139	0.706
ymggBS	M	Govav	1994	1.938	1.614	1.108	0.397	0.306	0.637	0.986	0.645
ymggGW	M	Govav	1994	2.037	1.635	1.106	0.248	0.224	0.597	1.036	0.549
ymggSY1	M	Govav	1985	1.973	1.603	1.217	0.302	0.265	0.700	1.136	0.568
ymggSY2	M	Govav	1991	1.913	1.597	1.142	0.249	0.377	0.737	1.117	0.660
ymmbSL	M	Bevdvn	1989	1.863	1.607	1.083	0.163	0.255	0.709	1.001	0.755

A.12 S1: Individual Speaker Durations

Table A.12: Speaker duration means per vowel (Wassink normalised)

Speaker	Sex	OriV	Yob	i	e	æ	u	o	a	ə	ɐ
mfagWA	F	Bevdvn	1970	134.53	226.51	182.12	137.45	180.84	159.48	91.07	97.46
mfbmHY	F	Bevdvn	1976	153.56	206.32	173.27	154.42	156.37	168.03	106.69	89.01
mfbmYS	F	Bevdvn	1972	173.87	192.51	172.21	149.35	145.87	183.90	90.54	88.12
mfgbAN	F	Govav	1970	150.34	223.13	189.25	151.77	183.86	165.12	79.95	72.39
mfggMW	F	Govav	1969	163.41	244.68	176.66	140.88	167.41	183.55	66.44	53.76
mfgmGwN	F	Govav	1965	151.83	198.66	178.66	141.83	144.83	181.62	104.17	99.73
mmbbBD	M	Bevdvn	1975	138.36	213.68	174.49	154.46	168.85	153.40	100.83	94.38
mmbbES	M	Bevdvn	1970	151.73	159.71	169.32	158.10	186.60	163.38	118.87	109.93
mmbbJS	M	Bevdvn	1962	146.95	198.41	167.84	162.38	163.24	172.83	91.13	94.42
mmbbKS	M	Bevdvn	1979	148.92	200.31	179.31	159.75	167.71	170.75	97.98	97.54
mmbbKS2	M	Bevdvn	1977	128.23	214.90	186.38	150.68	160.61	142.45	107.80	113.03
mmbbKS3	M	Bevdvn	1979	151.32	199.56	164.86	144.05	189.68	144.29	104.60	107.91
mmbbLM	M	Bevdvn	1975	144.61	184.22	189.61	160.57	166.25	159.20	98.03	100.57
mmbbLS	M	Bevdvn	1967	160.86	197.85	176.85	163.04	169.31	180.76	78.23	83.89
mmbbSS	M	Bevdvn	1980	150.59	180.05	171.85	142.91	173.32	164.93	108.59	110.71
mmbbST	M	Bevdvn	1971	154.90	176.40	183.12	155.26	173.92	148.04	106.72	108.41
mmbbTY	M	Bevdvn	1973	166.32	196.32	177.66	118.22	173.73	147.79	110.11	106.71
mmggDZ	M	Govav	1980	155.37	230.97	163.07	161.54	166.27	166.02	84.43	76.78
mmggGN	M	Govav	1975	154.54	186.30	174.80	149.55	164.66	185.95	99.89	92.59
mmggGS	M	Govav	1976	146.05	202.19	185.47	153.93	174.48	168.34	85.20	82.65
mmggGZ	M	Govav	1984	154.27	216.25	178.35	162.85	166.30	162.18	97.24	102.96
mmggMN	M	Govav	1968	138.18	210.90	204.70	153.00	182.96	165.37	98.06	93.76
mmggRG	M	Govav	1985	150.73	227.54	191.84	143.55	198.38	166.35	85.36	81.67
mmggRW	M	Govav	1989	149.33	169.51	180.85	153.25	163.75	192.26	106.18	90.52
mmggWW	M	Govav	1973	165.91	203.11	172.51	140.05	159.75	161.00	101.89	94.92
sfabKK	F	Bevdvn	1957	137.20	213.44	180.10	131.61	219.57	168.52	63.05	88.98
sfabWP	F	Bevdvn	1957	141.75	224.48	187.77	124.92	182.88	167.08	77.33	106.59
sfamRA	F	Bevdvn	1960	163.52	173.26	175.70	148.54	149.25	181.98	104.48	103.47
sfbaWZ	F	Bevdvn	1957	145.20	203.29	189.32	137.81	192.71	173.78	95.08	92.94
sfbbDS	F	Bevdvn	1957	156.97	196.55	199.35	165.82	182.70	136.15	88.97	81.05
sfbbKS	F	Bevdvn	1956	149.00	193.79	177.32	165.87	175.00	158.73	94.14	95.30
sfbgKS	F	Bevdvn	1953	144.14	186.42	175.00	154.08	190.02	165.82	98.66	96.63
sfgbMM	F	Govav	1949	166.34	212.99	165.25	146.48	152.86	204.44	63.16	62.48

Table A.12 – continued from previous page

Speaker	Sex	OriV	Yob	i	e	æ	u	o	a	ə	ɐ
sfgdWS	F	Govav	1968	144.94	195.40	179.93	154.22	191.41	145.85	95.34	85.09
sfggDS	F	Govav	1947	157.39	212.54	177.08	140.99	201.16	177.87	67.42	78.81
sfggZS	F	Govav	1949	160.54	202.06	173.33	152.17	169.14	184.22	81.85	85.37
sfgmFY	F	Govav	1955	148.42	193.97	179.14	162.60	132.45	182.28	93.16	88.99
smbbSZ	M	Bevdvn	1948	149.72	200.59	188.80	153.45	165.09	166.37	98.08	109.76
smbbTS	M	Bevdvn	1953	134.62	217.45	194.40	140.28	198.81	172.84	94.77	95.16
smbbYZ	M	Bevdvn	1951	152.55	200.59	181.10	155.06	171.55	167.34	104.11	106.80
smggGM	M	Govav	1954	122.39	223.65	207.93	120.88	172.64	177.33	91.93	100.82
smggMG	M	Govav	1954	161.59	195.62	194.39	156.35	200.12	162.88	69.09	78.70
smggWG	M	Govav	1947	127.51	230.60	198.75	146.41	193.72	169.56	84.22	89.20
smggWK	M	Govav	1943	159.23	209.78	182.68	160.14	184.36	159.31	79.74	78.05
smggZG	M	Govav	1949	162.49	198.37	169.02	152.50	183.23	184.63	93.03	83.08
smmgWW	M	Govav	1943	157.46	197.83	181.08	141.65	145.84	198.75	97.06	86.62
yfabRM	F	Bevdvn	1989	151.21	216.43	173.23	165.77	172.50	153.53	90.08	105.77
yfbbAT	F	Bevdvn	1989	145.88	181.79	177.83	138.31	167.62	151.52	109.22	119.22
yfbbJY	F	Bevdvn	1986	149.27	200.73	193.05	144.55	159.80	163.57	98.71	108.93
yfbbYL	F	Bevdvn	1991	140.20	221.61	185.31	155.64	171.54	164.66	91.44	95.96
yfggED	F	Govav	1987	133.71	205.23	188.64	137.81	176.64	141.95	105.88	113.98
yfggGY	F	Govav	1989	148.98	165.75	194.21	163.83	182.53	152.41	97.13	104.33
yfggJW	F	Govav	1985	133.43	217.16	180.64	137.43	189.21	148.83	98.86	107.83
yfggMW	F	Govav	1992	140.72	210.84	175.26	144.29	200.33	139.25	94.00	104.16
yfgmBZ	F	Govav	1990	149.09	189.98	163.93	161.69	160.25	188.48	94.09	95.11
ymbbMY	M	Bevdvn	1983	132.05	187.30	187.69	154.34	160.48	168.58	111.62	104.17
ymbbMZ	M	Bevdvn	1985	141.85	175.68	175.83	129.26	173.47	161.22	127.20	119.86
ymbbTT	M	Bevdvn	1982	161.93	176.08	182.99	164.89	178.37	157.43	96.43	103.82
ymggBS	M	Govav	1994	153.67	171.37	164.71	157.29	162.21	169.98	117.77	112.73
ymggGW	M	Govav	1994	146.49	183.82	188.35	148.95	173.05	175.82	103.64	114.81
ymggSY1	M	Govav	1985	149.58	173.05	186.93	133.56	188.20	143.48	117.02	116.43
ymggSY2	M	Govav	1991	143.42	175.15	202.72	128.28	213.93	147.34	103.38	102.78
ymmbSL	M	Bevdvn	1989	163.17	214.78	161.61	145.10	166.25	168.64	89.00	98.07

A.13 S1: Output of Statistical Modelling per Vowel

[e] F1						
Random	Var	SD				
Speaker	0.002	0.044				
Word	0.0211	0.145				
Residual	0.0247	0.157				
Fixed	Estimate	SE	df	t	p	
Intercept	-4.832	1.401	57.668	-3.448	0.001	**
YoB	0.003	0.001	57.407	4.034	0.000	***
SexM	0.044	0.022	60.189	1.999	0.050	
OriVGovav	0.014	0.022	64.039	0.625	0.534	

Table A.13: Model fit summary for [e] F1 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. Effect size: $R^2m = 0.05$, $R^2c = 0.51$. Observations = 305, 63 speakers, 8 words. Significance code: *** = p < .0001 - .001 ; ** = p < .001 - .01

[u] F1						
Random	Var	SD				
Speaker	0.008	0.093				
Word	0.002	0.044				
Residual	0.007	0.083				
Fixed	Estimate	SE	df	t	p	
(Intercept)	-3.228	1.690	61.752	-1.910	0.060	
YoB	0.002	0.001	61.725	2.046	0.045	*
OriVGov	-0.035	0.030	63.101	-1.348	0.182	
SexM	0.018	0.026	2.758	0.684	0.496	

Table A.14: Model fit summary for [u] F1 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. Effect size: $R^2m = 0.06$, $R^2c = 0.63$. Observations = 258, 63 speakers, 5 words. Significance code: * p < = .01 - .05

[æ] F1						
Random	Var	SD				
Speaker	0.003	0.051				
Word	0.016	0.127				
Residual	0.033	0.181				
Fixed	Estimate	SE	df	t	p	
Intercept	3.045	1.298	59.525	2.345	0.022	*
OriVGovav	0.063	0.020	63.047	3.120	0.003	**
YoB	-0.001	0.001	59.383	-1.282	0.205	
SexM	0.051	0.020	61.703	2.523	0.014	*

Table A.15: Model fit summary for [æ] F1 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. Effect size: $R^2m = 0.03$, $R^2c = 0.39$. Observations = 580, 63 speakers, 13 words. Significance code: ** = p < .001 - .01 ; * p < = .01 - .05

[æ] F2						
Random	Var	SD				
Speaker	0.003	0.051				
Word	0.016	0.127				
Residual	0.033	0.181				
Fixed	Estimate	SE	df	t	p	
Intercept	2.169	1.056	59.055	2.054	0.044	*
OriVGovav	-0.030	0.016	60.915	-1.847	0.070	
YoB	-0.001	0.001	58.963	-0.948	0.347	
SexM	0.029	0.016	60.029	1.745	0.086	

Table A.16: Model fit summary for [æ] F2 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. Effect size: $R^2m = 0.02$, $R^2c = 0.50$. Observations = 580, 63 speakers, 13 words. Significance code: * p < = .01 - .05

[o] F1						
Random	Var	SD				
Speaker	0.007	0.083				
Word	0.008	0.092				
Residual	0.020	0.143				
Fixed	Estimate	SE	df	t	p	
Intercept	5.098	1.606	62.233	-3.174	0.002	**
YoB	0.003	0.001	62.179	3.642	0.001	***
OriVGovav	-0.019	0.025	63.771	-0.788	0.434	
SexM	0.005	0.025	63.336	0.182	0.856	

Table A.17: Model fit summary for [o] F1 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. Effect size: $R^2m = 0.06$, $R^2c = 0.46$. Observations = 531, 63 speakers, 11 words. Significance code: *** = p < .0001 - .001 ; ** = p < .001 - .01

[o] F2						
Random	Var	SD				
Speaker	0.004	0.066				
Word	0.011	0.107				
Residual	0.007	0.083				
Fixed	Estimate	SE	df	t	p	
Intercept	-2.956	1.198	61.177	-2.467	0.016	*
YoB	0.002	0.001	61.080	2.636	0.011	*
OriVGovav	-0.025	0.018	62.072	-1.362	0.178	
SexM	0.053	0.019	61.739	2.857	0.006	**

Table A.18: Model fit summary for [o] F2 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. Effect size: $R^2m = 0.05$, $R^2c = 0.71$. Observations = 531, 63 speakers, 11 words. Significance code: ** = p < .001 - .01 ; * p < = .01 - .05

[ɑ] F1						
Random	Var	SD				
Speaker	0.003	0.055				
Word	0.006	0.074				
Residual	0.061	0.248				
Fixed	Estimate	SE	df	t	p	
Intercept	3.951	1.698	53.166	2.327	0.024	*
OriVGovav	-0.057	0.026	55.690	-2.178	0.034	*
YoB	-0.001	0.001	53.106	-1.364	0.178	
SexM	-0.011	0.026	54.995	-0.432	0.668	

Table A.19: Model fit summary for [ɑ] F1 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. Effect size: $R^2m = 0.02$, $R^2c = 0.14$. Observations = 533, 63 speakers, 10 words. Significance code: * p < = .01 - .05

[ɑ] F2	Var	SD				
Random	Var	SD				
Speaker	0.005	0.072				
Word	0.002	0.043				
Residual	0.007	0.083				
Fixed	Estimate	SE	df	t	p	
Intercept	1.179	1.290	63.265	0.914	0.364	
OriVGovav	-0.046	0.020	63.906	-2.335	0.023	*
YoB	-0.000	0.001	63.253	-0.374	0.710	
SexM	0.016	0.020	63.457	0.802	0.425	

Table A.20: Model fit summary for [ɑ] F2 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. Effect size: $R^2m = 0.04$, $R^2c = 0.53$. Observations = 533, 63 speakers, 10 words. Significance code: * p < = .01 - .05

[ə] F1						
Random	Var	SD				
Speaker	0.003	0.050				
Word	0.007	0.083				
Residual	0.031	0.177				
Fixed	Estimate	SE	df	t	p	
Intercept	0.596	1.369	52.676	0.436	0.665	
OriVGov	0.078	0.021	56.943	3.650	0.001	***
SexM	-0.023	0.022	56.218	-1.089	0.281	
Yob	0.000	0.001	52.564	0.180	0.858	

Table A.21: Model fit summary for [ə] F1 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. Effect size: $R^2m = 0.04$, $R^2c = 0.26$. Observations = 460, 63 speakers, 6 words. Significance code: *** = p < .0001 - .001

[ə] F2						
Random	Var	SD				
Speaker	0.003	0.057				
Word	0.014	0.119				
Residual	0.025	0.159				
Fixed	Estimate	SE	df	t	p	
Intercept	4.326	1.360	51.798	3.181	0.002	**
OriVGov	0.005	0.021	55.431	0.260	0.795	
Yob	-0.002	0.001	51.627	-2.353	0.022	*
SexM	0.018	0.021	54.833	0.862	0.393	

Table A.22: Model fit summary for [ə] F2 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. Effect size: $R^2m = 0.02$, $R^2c = 0.42$. Observations = 460, 63 speakers, 6 words. Significance code: ** = p < .001 - .01 ; * p < = .01 - .05

[ɐ] F2						
Random	Var	SD				
Speaker	0.002	0.048				
Word	0.014	0.120				
Residual	0.050	0.223				
Fixed	Estimate	SE	df	t	p	
Intercept	2.200	1.269	48.042	1.734	0.089	
SexM	-0.058	0.020	49.832	-2.948	0.005	**
OriVGovav	0.050	0.020	50.536	2.569	0.013	*
YoB	-0.001	0.001	47.909	-0.861	0.393	

Table A.23: Model fit summary for [ɐ] F1 with fixed effects of YoB, Gender, and VoO. Speaker and Word as random intercepts. . Effect size: $R^2m = 0.02$, $R^2c = 0.32$. Observations = 934, 63 speakers, 16 words. Significance code: ** = p < .001 - .01 ; * p < = .01 - .05

Bevdvn						
Random	Var	SD				
Speaker	0.004	0.063				
Word	0.015	0.122				
Residual	0.007	0.086				
Fixed	Coeff.	SE	df	t	p	
Intercept	-4.140	1.904	30.486	-2.594	0.014	*
SexM	0.039	0.025	31.778	1.537	0.134	
YoB	0.003	0.001	30.449	2.704	0.011	*

Govav						
Random	Var	SD				
Speaker	0.005	0.068				
Word	0.008	0.091				
Residual	0.006	0.075				
Fixed	Coeff.	SE	df	t	p	
Intercept	-1.872	1.522	24.495	-1.230	0.228	
SexM	0.066	0.027	29.780	2.456	0.020	*
YoB	0.001	0.001	29.474	1.343	0.190	

Table A.24: Model fit summary for Bevdvn and Govav separately: [o] F2 with fixed effects of YoB, and Gender. Speaker and Word as random intercepts. Bevdvn effect size: $R^2m = 0.06$, $R^2c = 0.73$. Govav: $R^2m = 0.07$, $R^2c = 0.72$. 293 observations, 32 speakers, 11 words. Significance code: * p < = .01 - .05

Bevdvn						
Random		Var	SD			
Speaker		0.002	0.047			
Word		0.017	0.129			
Residual		0.064	0.254			
Fixed	Coeff.	SE	df	t	p	
Intercept	1.554	2.112	23.928	0.736	0.469	
SexM	-0.135	0.029	27.830	-4.727	0.000	***
YoB	-0.000	0.001	23.853	-0.194	0.848	

Govav						
Random	Intercept	Var	SD			
Speaker		0.003	0.055			
Word		0.017	0.130			
Residual		0.053	0.230			
Fixed	Coeff.	SE	df	t	p	
Intercept	3.847	1.767	25.204	2.178	0.039	*
SexM	0.016	0.031	25.875	0.496	0.624	
YoB	-0.001	0.001	25.220	-1.550	0.134	

Table A.25: Model fit summary for Bevdvn and Govav separately: [ɐ] F1 with fixed effects of YoB, and Gender. Speaker and Word as random intercepts. Bevdvn: Speaker = 32, Word = 11, n = 293. Govav: Speaker = 31, Word = 11, n = 238. Significance codes: *** = p < .0001 - .001 ; * = .01

Random	Intercept	Var	SD			
Speaker		0.008	0.093			
Word		0.002	0.044			
Residual		0.007	0.083			
Fixed	Coeff.	SE	df	t	p	
(Intercept)	-3.228	1.690	61.752	-1.910	0.060	
YoB	0.002	0.001	61.725	2.046	0.045	*
OriV:Govav	-0.0350	0.030	63.101	-1.348	0.182	
Gender:M	0.018	0.0262	2.758	0.684	0.496	

Table A.26: Model fit summary for [ɐ] F1 with fixed effect of VoR, run on female speakers only. Speaker and Word as random intercepts. Speaker = 27, Word = 16, n = 1146. Significance codes: * p < = .01 - .05

A.14 S2: H-Drop Tokens by Speaker

Speaker	Sex	Yob	VoO	VoR	Retained (n, %)		Dropped (n, %)		Total (n)
sfggDS	F	1947	Govav	Govav	41	85%	7	15%	48
smbbSZ	M	1948	Bevdvn	Bevdvn	46	94%	3	6%	49
smggZG	M	1949	Govav	Govav	37	82%	8	18%	45
sfbfHZ	F	1949	Bevdvn	Bimadbn	8	57%	6	43%	14
smbbYZ	M	1951	Bevdvn	Bevdvn	70	93%	5	7%	75
smggMY	M	1952	Govav	Govav	22	56%	17	44%	39
smbbTS	M	1953	Bevdvn	Bevdvn	40	87%	6	13%	46
smggGM	M	1954	Govav	Govav	28	100%	0	0%	28
smggMG	M	1954	Govav	Govav	24	65%	13	35%	37
sfgmFY	F	1955	Govav	Bimadbn	30	52%	28	48%	58
sfbaWZ	F	1957	Bevdvn	Bevdvn	23	92%	2	8%	25
sfamRA	F	1960	Bevdvn	Bimadbn	44	72%	17	28%	61
mfgmGwN	F	1965	Govav	Bimadbn	16	35%	30	65%	46
mfgmGN	F	1966	Govav	Bimadbn	25	44%	32	56%	57
mmbbLS	M	1967	Bevdvn	Bevdvn	25	96%	1	4%	26
mmggMN	M	1968	Govav	Govav	13	22%	46	78%	59
mfaaRT	F	1969	Bevdvn	Bevdvn	34	74%	12	26%	46
mfggMW	F	1969	Govav	Govav	4	9%	41	91%	45
mmaaBA	M	1969	Bevdvn	Bevdvn	49	86%	8	14%	57
mfgbAN	F	1970	Govav	Bevdvn	29	49%	30	51%	59
mfbmYS	F	1972	Bevdvn	Bimadbn	34	60%	23	40%	57
mmggGS	M	1976	Govav	Govav	0	0%	13	100%	13
mmbbKS2	M	1977	Bevdvn	Bevdvn	25	61%	16	39%	41
mmbbKS	M	1979	Bevdvn	Bevdvn	5	10%	44	90%	49
ymbbTT	M	1982	Bevdvn	Bevdvn	0	0%	38	100%	38
mmggGZ	M	1984	Govav	Govav	8	24%	25	76%	33
mmggRG	M	1985	Govav	Govav	3	5%	55	95%	58
ymbbMZ	M	1985	Bevdvn	Bevdvn	0	0%	16	100%	16
yfbbJY	F	1986	Bevdvn	Bevdvn	1	2%	41	98%	42
yfbbAY	F	1987	Bevdvn	Bevdvn	1	3%	37	97%	38
yfabRM	F	1989	Bevdvn	Bevdvn	2	4%	43	96%	45
yfbbAT	F	1989	Bevdvn	Bevdvn	0	0%	35	100%	35
ymmbSL	M	1989	Bevdvn	Bevdvn	0	0%	28	100%	28
yfgmBZ	F	1990	Govav	Bimadbn	1	2%	52	98%	53
ymggSY2	M	1991	Govav	Govav	0	0%	6	100%	6
yfggMW	F	1992	Govav	Govav	1	3%	34	97%	35
ymggBS	M	1994	Govav	Govav	14	21%	54	79%	68
ymggGW	M	1994	Govav	Govav	3	5%	58	95%	61
ymbbME	M	1997	Bevdvn	Bevdvn	1	8%	12	92%	13
Total					707		942		1649

Table A.27: Token break down by speaker.

A.15 S2: H-Drop Token by Speaker, Organised by Gender

Female					Male				
Speaker	Ret.	Drop	YoB	VoO	Speaker	Ret.	Drop	YoB	VoO
sfggDS	41	7	1947	Govav	smbbSZ	46	3	1948	Bevdvn
sfbfHZ	8	6	1949	Bevdvn	smggZG	37	8	1949	Govav
sfgmFY	30	28	1955	Govav	smbbYZ	70	5	1951	Bevdvn
sfbaWZ	23	2	1957	Bevdvn	smggMY	22	17	1952	Govav
sfamRA	44	17	1960	Bevdvn	smbbTS	40	6	1953	Bevdvn
mfgmGwN	16	30	1965	Govav	smggGM	28	0	1954	Govav
mfgmGN	25	32	1966	Govav	smggMG	24	13	1954	Govav
mfaaRT	34	12	1969	Bevdvn	mmbbLS	25	1	1967	Bevdvn
mfggMW	4	41	1969	Govav	mmggMN	13	46	1968	Govav
mfgbAN	29	30	1970	Govav	mmaaBA	49	8	1969	Bevdvn
mfbmYS	34	23	1972	Bevdvn	mmggGS	0	13	1976	Govav
yfbbJY	1	41	1986	Bevdvn	mmbbKS2	25	16	1977	Bevdvn
yfbbAY	1	37	1987	Bevdvn	mmbbKS	5	44	1979	Bevdvn
yfabRM	2	43	1989	Bevdvn	ymbbTT	0	38	1982	Bevdvn
yfbbAT	0	35	1989	Bevdvn	mmggGZ	8	25	1984	Govav
yfgmBZ	1	52	1990	Govav	ymbbMZ	0	16	1985	Bevdvn
yfggMW	1	34	1992	Govav	mmggRG	3	55	1985	Govav
					ymmbSL	0	28	1989	Bevdvn
					ymggSY2	0	6	1991	Govav
					ymggBS	14	54	1994	Govav
					ymggGW	3	58	1994	Govav
					ymbbME	1	12	1997	Bevdvn
Total	294	470				413	472		

Table A.28: [h]-drop, Speakers organised by Gender. Ret. = [h]-form retained; Drop = [h]-dropped.

A.16 S2: All H-Words

Table A.29: All [h]-words from the NSC, ordered from most to least frequent.

Freq.	Word	Gloss	Retained	%	Dropped	%	n	%
1	hakr	boy	111	23.0%	372	77.0%	483	29.3%
2	hamba	village	124	50.6%	121	49.4%	245	14.9%
3	hkv	eye	89	43.0%	118	57.0%	207	12.6%
4	hambr	leg	55	28.5%	138	71.5%	193	11.7%
5	hi	bush.torch	111	70.7%	46	29.3%	157	9.5%
6	hrare	moon	56	50.9%	54	49.1%	110	6.7%
7	herge	life	68	70.8%	28	29.2%	96	5.8%
8	hure	straight	35	70.0%	15	30.0%	50	3.0%
9	hkvmbrto	prayer	6	25.0%	18	75.0%	24	1.5%
10	harq	thigh	19	82.6%	4	17.4%	23	1.4%
11	hkr	yam.cake	4	40.0%	6	60.0%	10	0.6%
12	hrnat	dry	6	66.7%	3	33.3%	9	0.5%
13	hañ	message	5	62.5%	3	37.5%	8	0.5%
14	hrr	belly	5	62.5%	3	37.5%	8	0.5%
15	hudu	walking.stick	1	12.5%	7	87.5%	8	0.5%
16	hrmḡar	upper.lip	6	100.0%	0	0.0%	6	0.4%
17	hara	bush.rope	4	80.0%	1	20.0%	5	0.3%
18	hrae	way	1	33.3%	2	66.7%	3	0.2%
19	havarama	leg.calf	1	50.0%	1	50.0%	2	0.1%
20	hkv-hkv	eyes	0	0.0%	2	100.0%	2	0.0%
			707		942		1649	

A.17 S2: Output of Statistical Modelling

Group	Variance	St. Dev.			
Speaker	3.02	1.74			
Word	2.76	1.66			
Fixed Effects	Coefficient	SE	Z Value	p	
Intercept	4.26	1.2	3.54	0.000	***
DISC	-1.74	1.12	-1.56	0.12	
GRM	0.51	1.46	0.35	0.73	
IU	-1.18	0.87	-1.36	0.17	
9MOD	-0.90	0.98	-0.92	0.36	
N	-1.32	0.93	-1.42	0.16	
POSS	-1.11	1.10	-1.01	0.31	
PRO	-0.57	1.14	-0.50	0.62	
V	-0.33	0.96	-0.35	0.73	
Village:Govav	-0.09	0.99	-0.09	0.93	

Table A.30: Logistic regression model fit summary on speakers aged under 45. [h]-drop YES vs NO, fixed effects of Preceding Word Class, and VoO. Speaker and Word as random intercepts. DEM = demonstratives (spatial and manner demonstratives); IU = final IU boundary marker; V = Verbs (inflected); DISC = discourse particles (topic markers, affirmatives; GRM = Grammatical Words (conjunctions, TAM particles, quotatives); MOD = Modifiers (numerals, quantifiers, temporal adverbials, negative particle, proper names, and nouns used attributively); N = Nouns (inflected, nominalised verbs); PRO = Pronouns (personal pronouns excluding the possessive form); POSS = possessives (possessive pronouns). Observations = 977, 21 speakers, 20 words. Significance codes: p < .000= '***', .01 = '*

R^2m = .02, R^2c = .44'

A.18 S3: All DEM+COPs

	Tokens (n)	Percentage	Person	Number	TAM
ge ym	339	66.0%	3	sg	nphd
ge tmaro	58	11.3%	3	sg	rmpst
ge ymaro	32	6.2%	3	sg	prim
ge wm	26	5.1%	1	sg	nphd
ge qmaro	14	2.7%	1	sg	rmpst
ge tm	10	1.9%	3	sg	ypst
ge wmaro	5	1.0%	1	sg	prim
ge yngm	5	1.0%	3	nsg	nphd
ge/gs em	4	0.8%	3	nsg	nphd
ge ere	3	0.6%	3	du	nphd
ge nm	3	0.6%	2	sg	nphd
ge tamaro	3	0.6%	3	nsg	rmpst
ge ynm	3	0.6%	1	nsg	nphd
ge qm	2	0.4%	1	sg	ypst
ge tnm	2	0.4%	1	nsg	ypst
ge tnmaro	2	0.4%	1	nsg	rmpst
ge tam	1	0.2%	3	nsg	ypst
ge tngm	1	0.2%	3	nsg	ypst
ge ynren	1	0.2%	1	du	nphd
Total	514	100.0%			

Table A.31: Summary of all DEM+COP form occurrences.

A.19 Other variables

Consultant Lawrence Mike commented that younger speakers use *bolo* 'old man' or 'poor fellow' in a certain way; a usage he finds to be an unpleasant innovation. - (section 3.3.1)

Degree of rounding in rounded bilabial stops. - (section 4.1.1.2)

Variable realisation of [z ~ ʒ ~ dʒ]. - (section 4.1.1.4)

Variable form of the singular ergative personal pronouns: *yndo* vs *yndon* (1sg), *bmo* vs *bmon* (2sg), *ymo* vs *ymon* (3sg). - (section 4.2.1.1).

The word for the numeral 6^2 *frta* vs *feta*. - (section 4.2.1.5).

The phonetic reduction of *fivi* 'body', 'real' as *vi*, occurring at the left edge of NPs. This may be part of the process by which the indefinite pronouns *emovi*, *nmbovi kanmavi*, and *nonavi* arose. - (section 4.2.1.7).

Variation in the form of the future marking particle *bä* vs *ä* - (section 4.3.1.1).

Occasional tendency for *de* to become phonologically merged with the prefix of the following main verb. If the prefix of the main verb is *y-* or *yn-* (third singular or first plural undergoer), the *de* particle leans in and often sounds like the prefix is *d-* or *dn-* - (section 4.3.1.2).

Variation in the negative particle *yao* vs *yawa*. *Yawa* appears to be derived by the phonetic merger of negative *yao* with a following *ä*, a reduced form of the future particle *bä*, i.e. *yao* NEG + *ä* FUT = yawa. - (section 4.3.3.1)

Allative and purposive/benefactive form and semantics overlap - (section 4.4.4.7, 4.4.4.5)

In naturalistic speech, there are words that sound like the adverbial demonstrative *gs* (section 4.3.3.2) followed by a *=va* producing the form *gs=va*. Consultants explained these instances as a phonetically reduced form of the particles *bva* and *dva* - (section 4.4.4.12

Variation in the form of the copulas where a final [n] is attached to the non-dual form, e.g. *ym* and *ymn* both 3sg nphd forms. Table 4.32 shows the copula forms with optional [n] where it has been observed in the corpus. There is no known meaning variation to the [n]-ful forms. - (section 4.5.8.1)

The ‘walk’ verb: My consultant Bunai Äniba rejected the form *ynd ynyutan* ‘We (pl.) walked’, and suggested the form *ynd ynyuta* without the final [n]. It is not clear whether his objection was to the final [n] in this particular verb context or for the entire paradigm, but it is a curious comment given that other prefixing verbs such as the copula and positionals show variation in realising word final [n]. Various speculations can be made as to whether this consultant’s judgement was a hyper-correction (i.e. the copula should not have a final [n], therefore any verb of a prefixing paradigm should not have [n]), or if he has some awareness of older forms that were possibly used by older speakers in the past (i.e. the ‘walk’ verb’s final *-tan* is a historical innovation)/

ɑ-set of prefixes used for verb with desinence of Perfective Past Habitual paradigm. - (section 4.7.0.6)

References

Abtahian, M. (2018). Style and language maintenance. [Presentation give at Sociolinguistic Symposium 22, Long Colliquium: Fresh insights on traditional variationist methods in non-English contexts, Auckland, New Zealand (June 28th 2018)].

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), 3099–3107.

Adank, P., van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America*, *116*(3), 1729–1738.

Adank, P., van Hout, R., & van de Velde, H. (2007). An acoustic description of the vowels of Northern and Southern Standard Dutch II. *The Journal of the Acoustical Society of America*, *121*(2), 1130–1141.

Aikhenvald, A. Y. (2002a). Traditional multilingualism and language endangerment. In D. Bradley & M. Bradley (Eds.), *Language endangerment and language maintenance: An active approach* (pp. 24–33). London, UK: Routledge.

Aikhenvald, A. Y. (2002b). Typological parameters for the study of clitics, with special reference to Tariana. In R. M. W. Dixon & A. Y. Aikhenvald (Eds.), *Word: A cross-linguistic typology* (pp. 42–78). Cambridge, UK: Cambridge University Press.

Aikhenvald, A. Y. (2008). *The Manambu Language of East Sepik, Papua New Guinea*. Oxford, UK: Oxford University Press.

Akimichi, T. (1998). Pig and man in Papuan societies: Two case from the Seltman of the fringe highlands and the Gidra of the lowland. *Senri Ethnological Studies*, 163–182.

Allen, J. (1982). Pre-contact trade in Papua New Guinea. In M. J. Ray &

H. Nelson (Eds.), *Melanesia: Beyond diversity, volume 1* (pp. 193–206). Canberra, Australia: Research School of Pacific Studies, The Australian National University.

Allen, J., & O'Connell, J. F. (2014). Both half right: Updating the evidence for dating first human arrivals in Sahul. *Australian Archaeology*, *79*, 86–108.

Anderson, S. R. (1992). *A-Morphous Morphology*. Cambridge, UK: Cambridge University Press.

Andrews, A. D. (2007). Relative Clauses. In T. Shopen (Ed.), *Language typology and syntactic description second edition: Volume 2: Complex constructions* (pp. 206–236). Cambridge, UK: Cambridge University Press.

Asiedu, R., & Sartie, A. (2010). Crops that feed the World 1 . Yams. *Food Security*, *2*, 305–315.

Astuti, R. (1995). "The Vezo Are Not a Kind of People " : Identity , Difference , and "Ethnicity " among a Fishing People of Western Madagascar. *American Ethnologist*, *22*(3), 464–482.

Athanasiadou, A., & Dirven, R. (1997). Conditionality, hypotheticality, counterfactuality. In A. Athanasiadou & R. Dirven (Eds.), *On conditionals again* (pp. 61–96). John Benjamins Publishing Company.

Atkinson, Q. D., & Gray, R. D. (2005). Curious Parallels and Curious Connections—Phylogenetic Thinking in Biology and Historical Linguistics QUENTIN. *Syst. Biol.*, *54*(4), 513–526.

Ayres, M. C. (1983). *This side, that side: Locality and exogamous group definition in Morehead area, southwestern Papua* (Unpublished PhD Thesis). University of Chicago.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390–412.

Baclawski, K. J. (2018). Diglossia and change from below in Eastern Cham. *Asia-Pacific Language Variation*, *4*(1), 73–102.

Ballard, C. (1994). Trade Networks and Sacred Geography in the Papua New Guinea Highlands. *Social Landscapes*, *29*(3), 130–148.

Baranowski, M. (2017). Class matters : the sociolinguistics of GOOSE and GOAT in Manchester English. *Language Variation and Change*, *29*, 301–339.

Barrau, J. (1968). L'Humide et al Sec. In A. P. Vayeda (Ed.), *Peoples and cultures of the pacific* (pp. 113–132). Garden City: Natural History Press.

Baxter, G. J., Blythe, R. A., Croft, W., & McKane, A. J. (2009). Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change*, *21*(2), 257–296.

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204.

Bell, A., & Holmes, J. (1992). H-droppin': Two sociolinguistic variables in New Zealand English. *Australian Journal of Linguistics*, *12*(2), 223–248.

Bercovitch, E. (1989). *Disclosure and concealment: a study of secrecy among the Nalumin people of Papua New Guinea* (Unpublished PhD Thesis). Stanford University.

Bickel, B. (2000). From Ergativus Absolutus to Topic Marking in Kiranti : A Typological Perspective. *Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society*, 38–49.

Billington, R., Thieberger, N., & Fletcher, J. (2018). Labial-velar consonants in Vanuatu : Evidence from Nafsan Labial-velar consonants :. In *Vanuatu languages conference 2018.* Port Vila, Vanuatu.

Blake, B. J. (1994). *Case.* Cambridge, UK: Cambridge University Press.

Blevins, J., Milin, P., & Ramscar, M. (2015). Zipfian discrimination. In V. Pirrelli, C. Marzi, & M. Ferro (Eds.), *Networds final conference.* Pisa.

Blevins, J., & Pawley, A. (2010). Typological implications of Kalam predictable vowels. *Phonology*, *27*(1), 1–44.

Blythe, R. A., & Croft, W. (2012a). S-curves and the mechanisms of propagation in language change. *Language*, *88*(2), 269–304. Retrieved from `http://muse.jhu.edu/content/crossref/journals/language/v088/88.2.blythe.html` doi: 10.1353/lan.2012.0027

Blythe, R. A., & Croft, W. (2012b). S-Curves and the mechanisms of propagation in language change. *Language*, 269–304.

Boersma, P., & Weenink, D. (2016). *Praat: Doing Phonetics by Computer.*

Book, M., Jankowski, B. L., Konnelly, L., & Tagliamonte, S. A. (2018). 'I don't come off as timid anymore ': Real-time change in early adulthood against the backdrop of the community. *Journal of Sociolinguistics*, 351–374.

Bourhis, R. Y., & Giles, H. (1977). The Language of Intergroup Distinctiveness. *Language, ethnicity and intergroup relations*, *13*, 119–135.

Bowern, C. (2010). Correlates of Language Change in Hunter-Gatherer and Other 'Small' Languages. *Language and Linguistics Compass*, *4*(8), 665–679.

Boyd, R. (2007). Cultural Adaptation and Maladaptation : Of Kayaks and Commissars. In S. W. Gangestad & J. A. Simpson (Eds.), *The evolution of mind : Fundamental questions and controversies* (pp. 327–331). New York, USA; London, UK: Guilford Publications.

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process.* Chicago: Chicago University Press.

Braunmüller, K., Höder, S., & Kühl, K. (2014). *Stability and Divergence in Language Contact: Factors and Mechanisms.* Amsterdam, and Philadelphia: John Benjamins Publishing Company.

Breban, T., Vanderbiesen, J., Davidse, K., Brems, L., & Mortelmans, T. (2012). Introduction. New reflections on the sources , outcomes , defining features and motivations of grammaticalization. In K. Davidse, T. Breban, L. Brems, & T. Mortelmans (Eds.), *Grammaticalization and language change: New reflections* (pp. 1–35). Amsterdam, and Philadelphia: John Benjamins Publishing Company.

Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and Language Change.* Cambridge, UK: Cambridge University Press.

Bromham, L., Hua, X., Fitzpatrick, T. G., & Greenhill, S. J. (2015). Rate of language evolution is affected by population size. *PNAS*, *112*(7), 2097–2102.

Brown, E. K., & Brown, E. L. (2012). Syllable-final and Syllable-initial /s/ reduction in Cali, Colombia: One variable or two? In R. J. File-Muriel & R. Orozco (Eds.), *Colombian varieties of spanish* (pp. 89–106). Iberoamericana, and Vervuert: Madrid, and Frankfurt.

Bucholtz, M. (1999). "Why be normal?": Language and identity practices in a community of nerd girls. *Language in Society*, *28*, 203–223.

Buchstaller, I. (2018). *(h) in Marshallese English.* Presentation given at Sociolinguistics Symposium 22, Sociolinguistic Variation in the Pacific Long Colloqia, Auckland, New Zealand (June 29th 2018).

Bybee, J. L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–85). Cambridge, UK: Cambridge University Press.

Bybee, J. L. (2010). *Language, Usage and Cognition.* Cambridge, UK: Cambridge University Press.

Bybee, J. L., & Hopper, P. J. (2001). Introduction to frequency and the emergence of linguistic structure. In *Frequency and the emergence of linguistic*

structure (pp. 1–24). John Benjamins Publishing Company.

Bybee, J. L., Perkins, R. D., & Pagliuca, W. (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World.* Chicago, USA: University of Chicago Press.

Bybee, J. L., & Thompson, S. A. (1997). Three frequency effects in syntax. *Berkeley Linguistic Society*, *23*, 65–85.

Caballero, G., & Harris, A. C. (2012). A working typology of multiple exponence. *Current Issues in Morphological Theory. Irregularities, analogy and frequency. Selected papers from the 14th International Morphology Meeting, Budapest (= Current Issues in Linguistic Theory 322)*(May 2010), 163–188.

Cahill, M. (1999). Aspects of the phonology of labial-velar stops. *Studies in African Linguistics*, *28*(2), 155–184.

Cahill, M. (2000). Positional contrast and labial-velars. *OSU Working Papers in Linguistics*, *53*, 71–92.

Campbell, L. (2001). What's wrong with grammaticalization? *Language Sciences*, *23*, 113–161.

Campbell, L., & Grondona, V. (2010). Who speaks what to whom ? Multilingualism and language choice in Misión La Paz. *Language in Society*, *39*(5), 617–646.

Campbell-Kibler, K. (2016). Towards a cognitively realistic model of meaningful sociolinguistic variation. In A. M. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 123–151). Cambridge, UK: Cambridge University Press.

Carroll, M. J. (2016). *The Ngkolmpu Language* (Unpublished PhD Thesis). Australian National University.

Carroll, M. J., Evans, N., Arka, I. W., Döhler, C., Kashima, E., Gast, V., ... Quinn, K. (2019). *Yamfinder: Southern New Guinea Lexical Database.*

Chafe, W. L. (1976). *The Caddoan, Iroquoian and Siouan languages* (Vol. 3.). The Hague: Mouton.

Chambers, J. K. (1992). Dialect Acquisition. *Linguistic Society of America*, *68*(4), 673–705.

Chappell, J. (2005). Geographic changes of coastal lowlands in the Papuan past. In A. Pawley, R. Attenborough, & J. Golson (Eds.), *Papuan pasts: Cultural, linguistic and biological histories of papuan-speaking peoples* (pp. 525–539). Canberra, Australia: Pacific Linguistics.

Chernela, J. M. (2013). Towards a Tukanoan Ethnolinguistics: Metadiscursive Practices, Identity, and Sustained Linguistic Diversity in the Vaupés Basin of Brazil and Colombia. In P. Epps & K. Stenzel (Eds.), *Upper rio negro: Cultural and linguistic interactio in northwestern amazonia* (pp. 197–244). Rio de Janeiro: Museu Naci.

Cheshire, J. (1982). *Variation in an English dialect: A sociolinguistic study.* Cambridge: Cambridge University Press.

Cheshire, J., & Fox, S. (2009). Was/were variation: A perspective from London. *Language Variation and Change*, *21*(1), 1–38.

Cheshire, J., Kerswill, P., Fox, S. P., & Torgersen, E. N. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics*, *15*(2), 151–196.

Childs, T., Good, J., & Mitchell, A. (2014). Beyond the Ancestral Code: Towards a Model for Sociolinguistic Language Documentation. *Language documentation and conservation*, *8*, 168–191.

Chirkova, K., Stanford, J. N., & Wang, D. (2018). A long way from New York City : Socially stratified contact-induced phonological convergence in Ganluo Ersu (Sichuan, China). *Language Variation and Change*, *30*, 109–145.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, 1–72.

Clarke, S. (2009). Sociolinguistic stratificiation and new dialect formation in a Canadian aboriginal community: not so different after all? In J. N. Stanford & D. R. Preston (Eds.), *Variation in indigenous minority langages* (pp. 109–128). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Clarkson, C., Jacobs, Z., Marwick, B., Fullagar, R., Wallis, L., Smith, M., ... Pardoe, C. (2017). Human occupation of northern Australia by 65,000 years ago. *Nature*, *547*(7663), 306–310.

Clopper, C. G., & Smiljanic, R. (2015). Regional variation in temporal organization in American English. *Journal of Phonetics*, *49*, 1–15.

Cohen, E. (2012). The Evolution of Tag-Based Cooperation in Humans. *Current Anthropology*, *53*(5), 588–616.

Comrie, B. (1976). *Aspect.* Cambridge, UK: Cambridge University Press.

Comrie, B. (1978). Ergativity. In W. P. Lehman (Ed.), *Syntactic typology:*

Studies in the phenomenology of language (pp. 329–394). Austin, USA: University of Texas Press.

Comrie, B. (2006). Syntactic typology: just how exotic ARE European-type relative clauses? In R. Mairal & J. Gil (Eds.), *Linguistic universals* (pp. 130–154). Cambridge, UK: Cambridge University Press.

Comrie, B., & Cysouw, M. (2012). New Guinea through the eyes of WALS. *Journal of the Linguistic Society of Papua New Guinea*, *1*(12), 65–94.

Conrad, R. (1978). Some notes on attitudes toward language and language choice in May River Iwam. In J. Murane (Ed.), *Language planning and grammatical typologies (working papers in papua new guinea languages 24)* (pp. 31–46). Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Cottet, F. (2014). Assibilation in Trans-New Guinea languages of the Bird's Head region. *Proceedings of the 2014 Speech Science & Technology Conference*, 172–175.

Cottet, F. (2015). *A Phonology of Mbahäm : Reduction & Contrast* (Unpublished doctoral dissertation). Australian National University.

Coupland, N. (1988). *Dialects in use: sociolinguistic variation in Cardiff English.* Cardiff: University of Wales Press.

Coupland, N. (2007). *Style: Language variation and identity.* Cambridge: Cambridge University Press.

Crisma, P. (2007). Were they 'dropping their aitches'? A quantitative study of h- loss in Middle English. *English Language and Linguistics*, *11*(1), 51–75.

Croft, W. (2000). *Explaining Language Change.* Harlow, UK: Pearson Education Limited.

Croft, W. (2006). The relevance of an evolutionary model to historical linguistics. In O. Nedergård Thomsen (Ed.), *Competing models of linguistic change: Evolution and beyond* (pp. 91–132). Amsterdam, Netherlands: John Benjamins Publishing Company.

Croft, W. (2008). Evolutionary linguistics. *Annual Review of Anthropology*, *37*(1), 219.

Cukor-Avila, P. (2000). Revisiting the observer's paradox. *American Speech*, *75*, 253–254.

Cukor-Avila, P. (2004). Language contact and the acquisition of AAVE: A case study of sociolectal adjustment. In *Language variation in the south iii (lavis iii).* Tuscaloosa, Alabama: University of Alabama.

Cukor-Avila, P., & Bailey, G. (2001). The effects of the race of the interviewer on sociolinguistic fieldwork. *Journal of Sociolinguistics*, *5*(2), 254–270.

Cukor-Avila, P., & Bailey, G. (2013). Real time and apparent time. In J. K. Chambers & N. Schilling-Estes (Eds.), *The handbook of variation and change* (pp. 239–262). John Wiley and Sons Inc.

Cummings, M. J. (2005). The role of Theme and Rheme in contrasting methods of organization in texts. In C. S. Butler, M. d. l. Á. G. Gómez, & S. M. Doval-Suárez (Eds.), *Topic* (pp. 129–154). John Benjamins Publishing Company. doi: 10.1075/pbns.140.12cum

Currie, T. E., & Mace, R. (2012). The Evolution of Ethnolinguistic Diversity. *Advances in Complex Systems*, *15*(01n02), 1150006.

Dahl, O. (2008). An exercise in a potseriori language sampling. *STUF - Language Typology and Universals Sprachtypologie und Universalienforschung*, *61*(3), 208–220.

Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological divesity: the linguistic niche hypothesis. *Advances in Complex Systems*, *15*(3/4), 1–16.

De Smet, H. (2016). How gradual change progresses: The interaction between convention and innovation. *Language Variation and Change*, *28*(01), 83–102.

De Vries, L. (1995). Demonstratives , referent identification and topicality in Wambon and some other Papuan languages. *Journal of Pragmatics*, *24*, 513–533.

De Vries, L. (2006). Areal pragmatics of New Guinea: Thematization, distribution and recapitulative linkage in Papuan narratives. *Journal of Pragmatics*, *38*, 811–828.

De Vries, L. (2012). Speaking of clans: Language in Awyu-Ndumut communities of Indonesian West Papua. *International Journal of the Sociology of Language*, *214*(214), 5–26.

Dediu, D., & Cysouw, M. (2013). Some Structural Aspects of Language Are More Stable than Others: A Comparison of Seven Methods. *PLoS ONE*, *8*(1).

Dediu, D., Cysouw, M., Levinson, S. C., Baronchelli, A., Christiansen, M. H., Croft, W., ... Lieven, E. (2013). Cultural Evolution of Language. In P. J. Richerson & M. H. Christiansen (Eds.), *Cultural evolution: Society, technology, language, and religion.* (pp. 303–331). Cambridge: MIT Press.

Degand, L., & Evers-Vermeul, J. (2015). Grammaticalization or pragmaticalization of discourse markers?: More than a terminological issue. *Journal of Historical Pragmatics*, *16*(1), 59–85.

de Lepercavche, M. (1967). Descent, Residence and Leadership in the New Guinea Highlands. *Oceania*, *52*(4), 320–343.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287–314.

Dench, A. (2001). Descent and diffusion: The complexity of the Pilbara situation. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics* (pp. 105–133). Oxford, UK: Oxford University Press. Retrieved from `e:{%}5CArticle{%}5CDench(2006-Descentanddiffusion).pdf`

Dench, A., & Evans, N. (1988). Multiple Case-Marking in Australian Languages. *Australian Journal of Linguistics*, *8*, 1–47.

Denes, P. B., & Pinson, E. N. (1963). *The Speech Chain: The physics and biology of spoken language* (2nd ed.). New York, USA: W.H. Freeman and Company.

Deshaies-Lafontaine, D. (1974). *A socio-phonetic study ofa Québec French community: Trois-Rivières* (Unpublished PhD thesis). University College London.

Deterding, D. (2001). Measurement of rhythm: a comparison of Singapore and British English. *Journal of Phonetics*(29), 217–230.

Diessel, H. (1999). *Demonstratives: Form, function, and grammaticalization.* Amsterdam, and Philadelphia: John Benjamins Publishing Company.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, *25*(2), 104–123.

Diewald, G. (2002). A model for relevant types of contexts in grammaticalization. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 103–120). John Benjamins Publishing Company.

Diewald, G., & Ferraresi, G. (2008). Semantic, syntactic and constructional restrictions in the diachronic rise of modal particles in German: A corpus-based study on the forma- tion of a grammaticalization channel. In E. Seoane & M. J. López-Couso (Eds.), *Theoretical and empirical issues in grammaticalization* (pp. 77–110). Amsterdam, Netherlands: John Benjamins Publishing Company.

Diewald, G., & Smirnova, E. (2010). *Evidentiality in German: Linguistic Real-*

ization and Regularities in Grammaticalization. Berlin, Germany: Mouton de Gruyter.

Dik, S. C. (1989). *The Theory of Functional Grammar*. Dordrecht: Foris.

Dingemanse, M., Torreira, F., & Enfield, N. J. (2013). Is "Huh?" a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS ONE*, *8*(11), 1–10.

Dinkin, A. J. (2008). The real effect of word frequency on phonetic variation: The real effect of word frequency on phonetic variation. *University of Pennsylvania Working Papers in Linguistics*, *14*(1), 97–106.

Dixon, R. M. W., & Aikhenvald, A. Y. (2004). *Adjective classes: A crosslinguistic typology*. Oxford, UK: Oxford University Press.

Döhler, C. (2016). *Komnzo : A language of Southern New Guinea* (Unpublished doctoral dissertation). Australian National University.

Döhler, C. (2018). *A grammar of Komnzo*. Berlin, Germany: Language Science Press.

Donohue, M. (1997). Tone Systems in New Guinea. *Linguistic Typology*, *1*(3), 347–386.

Donohue, M. (2008). Complexities with restricted numeral systems. *Linguistic Typology*, *12*, 423– 429.

Dorian, N. C. (1994). Varieties of variation in a very small place: Social homogeneity, prestige norms, and linguistic variation. *Language*, 631–696.

Dorian, N. C. (2010). *Investigating Variation : The Effects of Social Organization and Social Setting*. Oxford, UK: Oxford University Press.

Drager, K. K. (2015). *Linguistic variation, identity construction and cognition.*

Drager, K. K., & Hay, J. (2012). Exploiting random intercepts: Two case studies in sociophonetics. *Language Variation and Change*, *24*(01), 59–78.

Du Bois, J. W. (1987). The Discourse Basis of Ergativity. *Language*, *63*(4), 805–855.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of Discourse Transcription. *Talking Data: Transcription and Coding in Discourse Research*, 45–89.

Dubois, S., & Horvath, B. (1999). When the music changes, you change too: Gender and language change in Cajun English. *Language Variation and Change*, *11*(3), 287–313.

Eckert, P. (1989). The Whole Woman: Sex and gender difference in variation. *Language Variation and Change*, *1*(3), 245–267.

Eckert, P. (2004). Variation and a Sense of Place. In C. Fought (Ed.), *Sociolinguistic variation : Critical reflections* (pp. 107–118). Oxford, UK: Oxford University Press.

Eckert, P. (2008). Variation and the Indexical Field. *Journal of Sociolinguistics*, *12*(4), 453–476.

Eckert, P. (2012). Three Waves of Variation Study : The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*(41), 87–100.

Ellison, T. M., Evans, N., Kashima, E., Lindsey, K., Quinn, K., Schokkin, D., & Siegel, J. (2017). *Diachronic typology meets contact typology: a regional case study from Southern New Guinea.* [Presentation given at the Association of Linguistic Typology Conference, Canberra Australia (December 2017)].

Ellison, T. M., Evans, N., Kashima, E., Schokkin, D., & Williams, D. (2016). *Do shared norms cross language boundaries? A case study from multilingual Southern New Guinea.* Chayi, Taiwan.

Ellison, T. M., & Kashima, E. (2015). *Bilingualism and language divergence in small communities.* [Seminar Presentation at Hong Kong University, May 2015].

Ellison, T. M., & Miceli, L. (2017). Language monitoring in bilinguals as a mechanism for rapid lexical divergence. *Language*, *93*(2), 255–287.

Enfield, N. J. (2014). *Natural causes of language: frames, biases, and cultural transmission.* Berlin: Language and Science Press.

Epps, P. (2007). The Vaupés Melting Pot : Tucanoan Influence on Hup. In A. Y. Aikhenvald (Ed.), *Grammars in contact : A cross-linguistic typology* (pp. 267–289). Oxford, UK: Oxford University Press.

Epps, P. (2018). Contrasting linguistic ecologies: Indigenous and colonially mediated language contact in northwest Amazonia. *Language and Communication*, *62*, 156–169.

Epps, P., & Michael, L. (2017). The areal linguistics of Amazonia. In R. Hickey (Ed.), *The cambridge handbook of areal linguistics* (pp. 935–963). Cambridge, UK: Cambridge University Press.

Epps, P., & Salanova, A. P. (2013). The Languages of Amazonia. *Tipití: Journal of the Society for the Anthropology of Lowland South America*, *11*(1), 1–27.

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese.

The Journal of the Acoustical Society of America, *126*(3), 1379–1393.

Eshel, I. (1996). On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution. *Journal of Mathematical Biology*, *34*(5-6), 485–510.

Evans, N. (2003). Context, Culture, and Structuration in the Languages of Australia. *Annual Review of Anthropology*, *32*(1), 13–40.

Evans, N. (2009). Two pus one makes thirteen: Senary numerals in the Morehead-Maro region. *Linguistic Typology*, *13*, 321–335.

Evans, N. (2011). A tale of many tongues: documenting polyglot narrative in North Australian oral tradition. In B. Baker, I. Mushin, M. Harvey, & R. Gardener (Eds.), *Indigenous language and social identity. papers in honour of michael walsh* (pp. 291–314). Canberra, Australia: Pacific Linguistics.

Evans, N. (2012). Even more diverse than we had thought: The multiplicity of Trans-Fly languages. *Language Documentation & Conservation Special Publication No. 5*, 109–149.

Evans, N. (2013). Language Diversity as a Resource for Understanding Cultural Evolution. *Cultural Evolution: Society, Technology, Language, and Religion*, 233–268.

Evans, N. (2015a). Inflection in Nen. In M. Baerman (Ed.), *Oxford hanbook of inflection* (pp. 543–575). Oxford, UK: Oxford University Press.

Evans, N. (2015b). Restricted phonemes and interaction in Nen. [Workshop on the Languages of Melanesia, Kioloa Australia].

Evans, N. (2015c). Valency in Nen. In A. Malchukov & B. Comrie (Eds.), *Valency classes in the world's languages: Case studies from new guinea, australia, and the americas, and theoretical outlook* (pp. 1069–1116). Berlin, Germany: De Gruyter Mouton.

Evans, N. (2017a). *One thousand and one coconuts: growing memories in Southern New Guinea.* [Presentation given at European Society of Oceanicists Conference, Munich Germany June 2017].

Evans, N. (2017b). Quantification in Nen. In D. Paperno & E. L. Keenan (Eds.), *Handbook of quantifiers in natural language: Volume ii* (pp. 571–607). New York, USA: Springer.

Evans, N. (2018). The dynamics of language diversity. In R. Mesthrie & D. Bradley (Eds.), *Plenary and focus lectures from the 20th international congress of linguists* (pp. 11–41). Cape Town, South Africa: UCT Press.

Evans, N. (2019a). Coevolutionary Approaches to the Science of Language. In P. Pontarotti (Ed.), *Evolution, origin of life, concepts and methods* (pp. 195–213). Cham, Switzerland: Springer. doi: 10.1007/978-3-030-30363-1_10

Evans, N. (2019b). Linguistic divergence under contact. In M. Cennamo & C. Fabrizio (Eds.), *Historical linguistics 2015* (pp. 564–591). Amsterdam, The Netherlands: John Benjamins Publishing Company. doi: 10.1075/ijcl.18.1.09gri

Evans, N. (2019c). Waiting for the word: distributed deponency and the semantic interpretation of number in the Nen verb. In M. Baerman, A. Hippisley, & O. Bonds (Eds.), *Morphological perspectives. papers in honour of greville g. corbett.* (pp. 100–123). Edinburgh, UK: Edinburgh University Press.

Evans, N., Arka, I. W., Carroll, M. J., Choi, Y. J., Döhler, C., Gast, V., ... Siegel, J. (2018). The Languages of Southern New Guinea. In B. Palmer (Ed.), *The languages and linguistics of new guinea: A comprehensive guide* (pp. 640–774). De Gruyter Mouton.

Evans, N., Carroll, M. J., & Döhler, C. (2017). Proto-Yam Phonology Languages of Southern New Guinea. In *The languages of melanesia workshop.* Manokwari, Indonesia.

Evans, N., Ellison, T. M., & Kashima, E. (2018). Decategorialising prominence marker in Nen and Nmbo(PNG). [Presentation given at New Waves of Analysing Variation Asia Pacific 5, Brisbane Australia (February 1st 2018)].

Evans, N., & Klamer, M. (2012). Introduction: Linguistic challenges of the Papuan region. *Language Documentation & Conservation Special Publication No. 5*, *December 2*, 1–12.

Evans, N., & Miller, J. C. (2016). Nen. *Journal of the International Phonetic Association*, 1–19. doi: 10.1017/S0025100315000365

Evans, N., & Wilkins, D. (2000). In the Mind's Ear : The Semantic Extensions of Perception Verbs in Australian Languages. *Language*, *76*(3), 546–592.

Farr, C. J. M. (1999). *The interface between syntax and discourse in Korafe, a Papuan language of Papua New Guinea.* Canberra, Australia: Pacific Linguistics: Series C.

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialgoue of group monologue : The Influence of Group Size. *Psychological Science*, *11*(6), 481–486.

Fedden, S. (2011). *A grammar of Mian, a Papuan language of New Guinea.* Berlin, Germany: De Gruyter Mouton.

Fedden, S. (2012). Changes in Traditional Numerals Systems in Mian and other Trans New Guinea Languages. *Journal of the Lingusitic Society of Papua New Guinea*(Special Issue), 1–22.

Ferguson, C. A. (1990). From esses to aitches: identifying pathways of diachronic change. In W. Croft, K. Denning, & S. Kemmer (Eds.), *Studies in typology and diachrony for joseph h. greenberg* (pp. 59–78). Amsterdam: John Benjamins.

Ferree, S. (2000). Pouye. In J. Brownie (Ed.), *Sociolinguistic and literacy studies: South-west, sepik and morobe* (pp. 82–131). Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Fischer, O. (2011). Grammaticalization as analogically driven change? In B. Heine & H. Narrog (Eds.), *Oxford handbook of grammaticalization* (pp. 31–42). Oxford, UK: Oxford University Press.

Fishman, J. (1965). Varieties of ethnicity and varieties of language consciousness. *Georgetown Monographs on Language and Linguistics*, *18*, 69–79.

Fleischman, S. (1989). Temporal distance: a basic linguistic metaphor. *Studies in Language*, *13*(1), 1–50.

Foley, L. (1980). *Phonological Variation in Western Cherokee.* New York, USA; London, UK: Garland Publishing Inc.

Foley, W. (1986). *The Papuan Languages of New Guinea.* Cambridge, UK: Cambridge University Press.

Foley, W. (2000). The Languages of New Guinea. *Annual Review of Anthropology*, *29*, 357–404.

Foley, W. A. (2017). The morphosyntactic typology of Papuan languages. In *The languages and linguistics of the new guinea area* (pp. 895–938).

Forrest, J. (2017). The dynamic interaction between lexical and contextual frequency: A case study of (ING). *Language Variation and Change*, *29*(2), 129–156. doi: 10.1017/S0954394517000072

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409–438. doi: 10.1016/j.wocn.2005.08.002

Fox, B. A., & Thompson, S. A. (1990). A Discourse Explanation of the Grammar of Relative Clauses in English Conversation. *Linguistic Society of America*, *66*(2), 297–316.

Frajzyngier, Z. (2008). Grammaticalization, Typology and Semantics: Ex-

panding the Agenda. In M. J. López-Couso & E. Seoane (Eds.), *Rethinking grammaticalization: New perspectives* (pp. 61–102). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

François, A. (2012). The dynamics of linguistic diversity: Egalitarian multilingualism and power imbalance among northern Vanuatu languages. *International Journal of the Sociology of Language*, *214*(214), 85–110.

Fridland, V., Kendall, T., & Farrington, C. (2014). Durational and spectral differences in American English vowels: Dialect variation within and across regions. *The Journal of the Acoustical Society of America*, *136*(1), 341–349.

Gafter, R. J. (2016). What's a stigmatized variant doing in the word list? Authenticity in reading styles and Hebrew pharyngeals. *Journal of Sociolinguistics*, *20*(1), 31–58.

Giles, H., & Coupland, N. (1991). *Language: Contexts and consequences.* Belmot, USA: Thomson Brooks/Cole Publishing Co.

Giles, H., & Ogay, T. (2007). Communication Accommodation Theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 293–310). Mahwah, USA: Lawrence Erlbaum.

Gillespie, J. H. (1998). *Population Genetics: A concise guide.* Baltimore, USA: John Hopkins University Press.

Givón, T. (1983). Introduction. In T. Givón (Ed.), *Topic continuity in discourse: a quantitative cross-language study* (pp. 11–42). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Godelier, M. (1986). *The making of great men.* Cambridge, UK: Cambridge University Press.

Gomez-Imbert, E. (1996). When Animals Become "Rounded" and "Feminine": Conceptual Categories and Linguistic Classification in a Multilingual Setting. In *Rethinking linguistic relativity* (pp. 438–469).

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: Its Origins and Evolution.* Cambridge, UK and Massachusetts, USA: Cambridge University Press.

Grace, G. W. (1975). *Linguistic diversity in the Pacific: On the sources of diversity.* Vancouver, Canada.

Grace, G. W. (1981). History of Research in Austronesian Languages of the New Guinea Area: General. In J. Lynch (Ed.), *Readings in the comparative*

linguistics of melanesia. studying pacific language series, no. 4 (pp. 69–77). Papua New Guinea: Language Department, University of Papua New Guinea.

Grama, J. (2015). *Variation and Change in Hawai'i Creole Vowels* (Unpublished doctoral dissertation). University of Hawai'i.

Greenhill, S. J. (2014). Demographic correlates of language diversity. In *Routledge handbook of historical linguistics* (pp. 557–578).

Greenhill, S. J. (2015). *Evolution and Language: Phylogenetic Analyses* (Vol. 8). Elsevier. doi: 10.1016/B978-0-12-384719-5.00102-7

Greenhill, S. J., Wu, C.-H., Hua, X., Dunn, M., Levinson, S. C., & Gray, R. D. (2017). Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*, 201700388.

Grosjean, F. (2001). The Bilingual's Language Modes. In J. Nichol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1–22). Malden, USA: Wiley-Blackwell.

Gundel, J. K., & Fretheim, T. (2008). Topic and Focus. *The Handbook of Pragmatics*, 175–196.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge, UK: Cambridge University Press.

Guy, G. R. (1980). Variation in the group and the individual: the case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1–36). New York, USA; London, UK: Academic Press.

Haiman, J. (1978). Conditionals are Topics. *Language*, *54*(3), 564–589.

Hale, K. (1976). The adjoined relative clause in Australia. In R. M. W. Dixon (Ed.), *Grammatical categories in australian languages* (pp. 78–105). Canberra, Australia: Australian Institute of Aboriginal Studies.

Hammarström, H. (2009). Whence the Kanum base-6 numeral system? *Linguistic Typology*, *13*, 305–319.

Hammarström, H., & Nordhoff, S. (2012). The languages of Melanesia: Quantifying the level of coverage. *Language Documentation & Conservation Special Publication No. 5*, *December 2*, 13–33.

Harris, A. C., & Campbell, L. (1995). *Historical Syntax in Cross-Linguistic Perspective*. Cambridge, UK: Cambridge University Press.

Hasan, R., & Fries, P. H. (1995). Reflections on subject and theme: An introduction. (1995), xiii. doi: 10.1075/cilt.118.01has

Haspelmath, M. (1997). *Indefinite Pronouns*. Oxford, UK: Oxford University

Press.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, *86*(3), 663–687.

Haspelmath, M. (2011). On S , A , P , T , and R as comparative concepts for alignment typology. *Linguistic Typology*, *15*(3), 1–26.

Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (2008). *The world atlas of language structures online*. Munich: Max Planck Digital Library.

Haudricourt, A. G. (1961). Rihesse en phonémes et cichesse en locuteurs. *L'Homme*, *1*, 5–10.

Haugen, E. (1972). The Ecology of Language. In A. S. Dil (Ed.), *The ecology of language: Essays by einar haugen* (pp. 325–339). Stanford, USA: Stanford University Press.

Hay, J., & Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. *Language*, *92*(2), 298–330.

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83–91.

Hays, T. E. (2005). Vernacular names for tubers in Irian Jaya: implications for agricultural history. In *Papuan pasts: Cultural, linguistic and biological histories of papuan-speaking peoples* (pp. 625–672). Canberra, Australia: Pacific Linguistics.

Hazen, K. (2002). Identity and Language Variation in a Rural Community. *Language*, *78*(2), 240–257.

Heeschen, V. (1994). How long are clauses and sentences in a Papuan language like Eipo? In G. P. Reesink (Ed.), *Topics in descriptive papuan linguistics* (pp. 50–74). Leiden, The Netherlands: (Semaian 10). Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden.

Heeschen, V. (1998). *An ethnographic grammar of the Eipo language. (Mensch, Kultur und Umwelt im Zentralen Bergland von West-Neuguinea, 23.)* (Vol. 53) (No. 9). Berlin, Germany: Dietrich Reimer. doi: 10.1017/CBO9781107415324.004

Heffernan, K. (2010). Mumbling is macho: Phonetic distinctiveness in the speech of American radio DJs. *American Speech*, *85*(1), 67–90. doi: 10.1215/00031283-2010-003

Heine, B. (2002). On the role of context in grammaticalization. *New Reflections on Grammaticalization*, 83–101.

Heine, B. (2003). Grammaticalization. In B. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 575–599). Malden, USA; Oxford, UK: Blackwell Publishing.

Heine, B., & Kuteva, T. (2002a). On the Evolution of Grammatical Forms. In A. Wray (Ed.), *The transition to language* (Vol. 2, pp. 376–397). Oxford, UK: Oxford University Press.

Heine, B., & Kuteva, T. (2002b). *World Lexicon of Grammaticalization*. Cambridge, UK: Cambridge University Press.

Henderson, J. (1995). *Phonology and Grammar of Yele, Papua New Guinea* (Vol. Series B-1). Canberra, Australia: Pacific Linguistics.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behav Brain Sci*, *33*(2-3), 61–135.

Henton, C. (1992). Acoustic variability in the vowels of female and male speakers. *The Journal of the Acoustical Society of America*, *91*(4), 2387–2387. doi: 10.1121/1.403313

Henton, C. G. (1985). *A Comparative Study of Phonetic Sex-specific Differences Across Languages* (Doctor of Philosophy). University of Oxford.

Hickey, R. (2017). Areas, areal features, and areality. In R. Hickey (Ed.), *The cambridge handbook of areal linguistics* (pp. 1–16). Cambridge, UK: Cambridge University Press.

Hildebrandt, K. A., Jany, C., & Silva, W. (2017). Documenting variation in endangered languages. *Language Documentation & Conservation special publication*, *13*, 1–5.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Hinds, J. (1983). Topic continuity in Japanese. In T. Givón (Ed.), *Topic continuity in discourse: a quantitative cross-language study* (pp. 43–94). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Hitchcock, G. (2009). William Dammköhler's third encounter with the tugeri (Marind-Anim): a Dutch expedition of exploration and punishment in south New Guinea. *The Journal of Pacific History*, *44*(1).

Hitchcock, G. (2010). Mound-and-ditch taro gardens of the Bensbach or Torassi River area, southwest Papua New Guinea. *The Artefact*, *33*, 70–90.

Hoffman, M. F., & Walker, J. A. (2010). Ethnolects and the city: Ethnic

orientation and linguistic variation in Toronto English. *Language Variation and Change*, *22*(01), 37–67.

Honeyman, T. (2016). *A grammar of Momu: A language of Papua New Guinea* (Unpublished PhD Thesis). Australian National University.

Hopper, P. J. (1991). On some principles of grammaticization. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaicalization, volume 1* (pp. 17–35). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Hopper, P. J., & Traugott, E. C. (1993). *Grammaticalisation*. Cambridge, United Kingdom: Cambridge University Press.

Horvath, B. (1985). *Variation in Australian English: The Sociolects of Sydney*. London, UK: Cambridge Studies in Linguistics.

Houston, A. C. (1985). *Continuity and change in English morphology: The variable (ING)* (Unpublished PhD Thesis). University of Pennsylvania.

Hua, X., Greenhill, S. J., Cardillo, M., Schneemann, H., & Bromham, L. (2019). The ecological drivers of variation in global language diversity. *Nature Communications*, 1–10.

Hull, D. L. (1988). *Science as a process: an evolutionary accoutn of the social and conceptual development of socience*. Chicago, USA: University of Chicago Press.

Irvine, A. (2004). A good command of the English language: Phonological variation in the Jamaican acrolect. *Journal of Pidgin and Creole Languages*, *19*(1), 41–76.

Jacewicz, E., Fox, R. A., & Salmons, J. (2011). Cross-generational vowel change in American English. *Language Variation and Change*, *23*(1), 45–86.

Jackson, J. (1974). Language identity of the Colombian Vaupes Indians. In R. Bauman & J. Sherzer (Eds.), *Explorations in the ethnography of speaking* (pp. 50–64). Cambridge, UK: Cambridge University Press.

Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Linguistics and Language Compass*, *3*(1), 359–383.

Johnson, K. (2006). Resonance in an exemplar-based lexicon : The emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499. doi: 10.1016/j.wocn.2005.08.004

Johnson, K., & Martin, J. (2001). Acoustic Vowel Reduction in Creek : Effects of. *Phonetica*, *58*, 81–102.

Johnstone, B. (2011). Language and Place. In R. Mesthrie (Ed.), *Cambridge*

handbook of sociolinguistics (pp. 203–218). Cambridge, UK: Cambridge University Press. doi: 10.2307/j.ctt7sjpn.8

Jones, G. M. (2014). Secrecy. *Annual Review of Anthropology*, *43*, 53–69.

Kashima, E., Williams, D., Ellison, T. M., Schokkin, D., & Escudero, P. (2016). Uncovering the acoustic vowel space of a previously undescribed language: the vowels of Nambo. *Journal of the Acoustical Society of America Express Letters*, *139*(6).

Keesing, R. M. (1991). Substrates, calquing and grammaticalization in Melanesian Pidgin. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization, volume 1* (pp. 315–342).

Kendall, A. T., Thomas, E. R., & Kendall, M. T. (2018). *Package 'vowels ' (Version 1.2-2).* Retrieved from `http://blogs.uoregon.edu/vowels/`

Kirchner, R. (2004). Consonant lenition. *Phonetically Based Phonology*(2004), 314–345.

Kisler, T., Schiel, F., & Sloetjes, H. (2012). Signal processing via web services: the use case WebMAUS. *In Proceedings Digital Humanities 2012*, 30–34.

Klamer, M. (2004). Multi-categoria items as underspecified lexical entries: the case of Kambera waéngu. In O. Fischer, M. Norde, & H. Perridon (Eds.), *Up and down the cline - the nature of grammaticalization* (pp. 299–324). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Klamer, M. (2018). Typology and grammaticalization in the Papuan languages of Timor, Alor and Pantar. In H. Narrog & P. Pardeshi (Eds.), *Grammaticalization from a typological perspective* (pp. 235–460). Oxford, UK: Oxford University Press.

Kroskrity, P. V. (2018). On recognizing persistence in the Indigenous language ideologies of multilingualism in two Native American Communities. *Language and Communication*, *62*, 133–144.

Kulick, D. (1992). Anger, gender, language shift and the politics of revelation in a Papua New Guinean Village. *Pragmatics*, *2*(3), 281–296.

Kulick, D. (1997). *Language shift and cultural reproduction: socialization, self, and syncretism in a Papua New Guinean village.* Cambridge: Cambridge University Press.

Kuno, S. (1973). *The Structure of the Japanese Language.* Cambridge, Massachusetts, USA: MIT Press.

Kuryłowicz, J. (1965). the evolution of grammatical categories. *Diogenes*, *51*,

55–71.

Kuster, W. (2003). *Linguistic Complexity: The Influence of Social Change on Verbal Inflection.* Leiden, The Netherlands: Royal University of Leiden Center for Linguistics.

Labov, W. (1965). On the mechanisms of linguistic change. *Georgetown Monographs on Language and Linguistics*, *18*, 91–114.

Labov, W. (1966). *The Social Stratification of English in New York City.* Washington, DC: Centre for Applied Linguistics.

Labov, W. (1972a). Negative attraction and negative concord in English grammar. *Language*, *48*, 773–818.

Labov, W. (1972b). The social motivations of sound change. In *Sociolinguistic patterns* (pp. 1–42). Philadelphia: University of Pennsylvania Press.

Labov, W. (1972c). *Sociolinguitic Patterns.* Philadelphia: University of Pennsylvania Press.

Labov, W. (1989). The Child As Linguistic Historian. *Language Variation And Change*, *1*(1), 85–97. doi: 10.1017/S0954394500000120

Labov, W. (1994). *Principles of Linguistic Change, Volume I: Internal Factors.* Oxford, UK: Wiley-Blackwell.

Labov, W. (2001). *Principles of Linguistic Change, Vol II: Social Factors.* Oxford, UK: Blackwell.

Labov, W. (2007). Transmission and Diffusion. *Language*, *83*(2).

Labov, W., Ash, S., & Bober, C. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change.* Berlin, Germany: De Gruyter Mouton.

Lacerda, F. (1995). The perceptual-magnet effect: An emergent consequence of exemplar based phonetic memory. In K. Elenius & P. Branderyd (Eds.), *Proceesings of the xiiith international congress of phonetic sciences* (Vol. 2, pp. 140–147). Stockholm.

Laufer, A. (1991). The 'glottal fricatives'. *Journal of the International Phonetic Association*, *21*(2), 91–93.

Lavandera, B. R. (1978). Where Does the Sociolinguistic Variable Stop? *Language and Society*, *7*(2), 171–182.

Laycock, D. (1979). Multilingualism: linguistic boundaries and unsolved problems in Papua New Guinea. In S. A. Wurm (Ed.), *New guinea and neighboring areas: A sociolinguistic laboratory* (pp. 81–100). The Hague, Netherlands: Mouton.

Laycock, D. C. (1982). Melanesian linguistic diversity: a Melanesian choice? In R. J. May & H. Nelson (Eds.), *Melanesia: Beyond diversity, volume 1* (pp. 33–38). Canberra, Australia: Research School of Pacific Studies, The Australian National University.

Laycock, D. C. (1991). Linguistic Diversity in Melanesia: A tentative explanation. In R. Carle, M. Heinschke, P. W. Pink, C. Rost, & K. Stradtlander (Eds.), *Gava: Studies in austronesian languages and cultures* (pp. 31–37). Berlin: Reimer.

Laycock, D. C. (2003). *A dictionary of Buin: A language of Bougainville.* Canberra, Australia: Pacific Linguistics.

Le Page, R. B., & Tabouret-Keller, A. (1985). *Acts of identity: Creole-based approaches to language and ethnicity.* Cambridge, UK: Cambridge University Press.

Lebot, V. (2009). *Tropical Root and Tuber Crops: Cassava, Sweet Potato, Yams and Aroids.* Oxforshire UK, Cambridge USA: CABI.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English based on the British national corpus.* London, UK: Longman.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, Massachusetts, USA: MIT Press.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*(JUN), 1–17.

Lichtenberk, F. (1991). On The Gradualness of Grammaticalisation. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization volume i. theoretical and methodological issues* (pp. 37–80). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Lindsey, K. (2019). *Ghost Elements in Ende Phonology* (Unpublished PhD Thesis). Stanford University.

Lindström, E., & Remijsen, B. (2005). Aspects of the prosody of Kuot , a language where intonation ignores stress. *Linguistics*, *43*(4), 839–870.

Lippi-Green, R. L. (1989). Social network integration and language change in progress in a rural alpine village. *Language in society*, *18*(2), 213–234.

Litteral, R. (1978). Changes in the Bibriari communicative system. In J. Murane (Ed.), *Language planning and grammatical typologies (working papers in papua new guinea languages 24)* (pp. 25–30). Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Llamas, C. (2007). "A place between places": Language and identities in a border town. *Language in Society*, *36*(4), 579–604. doi: 10.1017/S0047404507070455

Locke, J. L. (1993). *The child's path to spoken language*. Harvard University Press.

Loughnane, R. (2009). A grammar of Oksapmin. *University of Melbourne dissertation*.

Lüpke, F. (2016). Uncovering Small-Scale Multilingualism. *Critical Multilingualism Studies*, *4*(2), 35–74.

Lüpke, F. (2018). Multiple Choice: Language use and cultural practice in rural Casamance between convergence and divergence. In J. Knörr & W. T. Filho (Eds.), *Creolization and pidginisation in contexts of post-colonial diversity: Language, culture, and identity* (pp. 181–208).

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, *5*(1).

Lynch, S. (2016). Don't Put Your 'and on the Hoven: A variationist study of /h/ insertion and /h/ dropping in Kosraean English..

Macintyre, M., & Young, M. (1982). The persistence of traditional trade and ceremonial exchange in the Massim. In M. J. Ray & H. Nelson (Eds.), *Melanesia: Beyond diversity, volume 1* (pp. 207–222). Canberra, Australia: Research School of Pacific Studies, The Australian National University.

Maclagan, M., & Hay, J. (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English "short" front vowel shift. *Language Variation and Change*, *19*(1), 1–25.

Maddieson, I. (2013). Vowel Quality Inventories. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology.

Mair, C. (2004). Corpus linguistics and grammaticalisation theory: Statistics, frequencies, and beyond. In H. Lindquist & C. Mair (Eds.), *Corpus approaches to grammaticalization in english* (pp. 121–150). John Benjamins Publishing Company.

Mansfield, J. B. (2015). Consonant lenition as a sociophonetic variable in Murrinh Patha (Australia). *Language Variation and Change*, *27*(2), 203–225.

Mansfield, J. B., & Stanford, J. N. (2017). Documenting Sociolinguistic Variation in Lesser- studied Indigenous Communities: Challenges and Practical Solutions. *Language Documentation & Conservation*, *13*, 116–136.

Martin, G. C. (2001). *A study of time as being according to the Keraakie people of Southwest Papua New Guinea* (Unpublished doctoral dissertation). The University of Sydney, Sydney, Australia.

Martin, J. B. (2010). How to Tell a Creek Story in Five Past Tenses 1. *International Journal of American Linguistics*, *76*(1), 43–70.

Martin, P. (1998). Les voyelles d'aperture moyenne en français du Québec. *Cahiers de l'ILSL*, *11*, 215–242.

McDougall, I., Brown, F. H., & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature*, *433*(7027), 733–736.

McElhannon, K. A. (1971). Classifying New Guinea Languages. *Anthropos*, *66*(1-2), 120–144.

Meek, B. A. (2007). Respecting the Language of Elders: Ideological Shift and Linguistic Discontinuity in a Northern Athapascan Community. *Journal of Linguistic Anthropology*, *17*(1), 23–43. doi: 10.1525/jlin.2007.17.1.23

Meggitt, M. J. (1974). "Pigs Are Our Hearts!" The Te Exchange Cycle among the Mae Enga of New Guinea. *Oceania*, *44*(3), 165–203.

Meillet, A. (1912). L'évolution des formes grammaticales. *Scientia*, *12*(6).

Mendoza-Denton, N. (2002). Language and identity. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 475–499). Malden, USA: Blackwell Publishing.

Mendoza-Denton, N. (2008). *Homegirls: Language and Cultural Practice Among Latina Youth Gangs*. Malden, USA; Oxford, UK: Blackwell Publishing.

Merlan, F. (1981). Land, Language and Social Identity in Aboriginal Australia. *Mankind*, *13*(2), 133–148.

Merlan, F. (1997). The Mother-in-law Taboo: Avoidance and Obligation in Aboriginal Australian Society. In F. Merlan, J. Morton, & A. Rumsey (Eds.), *Scholar and skeptic: Australian aboriginal studies in honour of lr hiatt* (pp. 95–122).

Mesthrie, R. (2010). Socio-phonetics and social change: Deracialisation of the GOOSE vowel in South African English. *Journal of Sociolinguistics*, *14*(1), 3–33.

Meyerhoff, M. (1999). Sorry in the Pacific : Defining Communities , Defining Practice. *Language and Society*, *28*(2), 225–238.

Meyerhoff, M. (2017). Writing a linguistic symphony: Analyzing variation while doing language documentation. *Canadian Journal of Linguistics/Revue*

canadienne de linguistique, *62*(04), 525–549.

Meyerhoff, M., & Holmes, J. (1999). The Community of Practice: Theories and methodologies in language and gender research. *Language in Society*, *28*(2), 173–183.

Meyerhoff, M., & Klaere, S. (2017). A case for clustering speakers and linguistic variables: Big issues with smaller samples in language variation. In I. Buchstaller & B. Siebenhaar (Eds.), *Language variation – european perspectives vi: Selected papers from the eighth international conference on language variation in europe (iclave 8), leipzig, may 2015 (studies in language variation 19)* (pp. 23–46). Amsterdam, Netherlands: John Benjamins Publishing Company.

Meyerhoff, M., & Stanford, J. N. (2015). The changing face of sociolinguistics with a global perspective. In D. Smakman & P. Heinrich (Eds.), *Globalising sociolinguistics.* New York, USA: Routledge.

Meyerhoff, M., & Walker, J. A. (2012). Grammatical variation in Bequia (St Vincent and the Grenadines). *Journal of Pidgin and Creole Languages*, *27*(2), 209–234.

Michael, L. (2014). Social Dimensions of Language Change. In B. Evans (Ed.), *Routledge handbook of historical linguistics* (pp. 484–502).

Miller, J. C. (2013). *The Phonetics of Tone in Two Dialects of Dane-zaa (Athabaskan)* (Unpublished doctoral dissertation). University of Washington, Washington, USA.

Milroy, J. (1992). *Linguistic variation and change: on the historical sociolinguistics of English.* Oxford, UK: Basil Blackwell.

Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics*, *21*(2), 339–384.

Milroy, L. (1980). *Language and social networks.* Oxford, UK: Blackwell.

Milroy, L. (1987). *Observing and analysing natural language: A critical account of sociolinguistic method* (1992nd ed.). Oxford, UK: Wiley-Blackwell.

Mitchell, A. (2015). *Linguistic avoidance and social relations in Datooga* (Unpublished PhD Thesis). University at Buffalo, State University of New York.

Mooi, E., & Sarstedt, M. (2011). Cluster Analysis. In E. Mooi & SarstedM (Eds.), *A concise guide to market research* (pp. 273–324). Berlin, Germany: Springer-Verlag.

Moore, E., & Carter, P. (2015). Dialect contact and distinctiveness: The so-

cial meaning of language variation in an island community. *Journal of Sociolinguistics*, *19*, 3–36.

Moraes, J. A., Callou, D., & Leite, Y. (1996). O sistema vocálico do português do Brasil: caracterização acústica (the vocalic system of the Portuguese of Brazil: Acoustic characterization). In M. Kato (Ed.), *Gramática do português falado (the grammar of spoken portuguese)* (pp. 33–53). Editora da Unicamp, Campinas.

Mosel, U. (2006). Sketch grammar. In J. Gippert, N. P. Himmelmann, & U. Mosel (Eds.), *Essentials of language documentation* (pp. 301–309). Berlin, Germany ; New York, USA: Mouton de Gruyter.

Mous, M. (1994). Ma'a or Mbugwe. In D. Bakker & M. Mous (Eds.), *Mixed languages* (pp. 175–200). Amsterdam, The Netherlands: Institute for Functional Research into Language and Language Use (IFOTT).

MPI. (2018). *ELAN (Version 5.2).* Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics, The Language Archive.

Mufwene, S. S. (2001). *The Ecology of Language Evolution.* Cambridge: Cambridge University Press.

Mugglestone, L. (2003). / h / and Other Symbols of the Social Divide. In L. Mugglestone (Ed.), *Talking proper : The rise of accent as social symbol.* (2nd ed., pp. 95–134). Oxford, UK: Oxford University Press.

Mühlhäusler, P. (2006). Naming Languages, Drawing Language Boundaries and Maintaining Languages with Special Reference to the Linguistic Situation in Papua New Guinea. In D. Cunningham, D. E. Ingram, & K. Sumbuk (Eds.), *Language diversity in the pacific : endangerment and survival* (pp. 24–39). Clevedon, UK ; Buffalo, USA: Multilingual Matters. doi: 10.21832/9781853598685-006

Muradoğlu, S. (2017). *When is enough enough ? A corpus-based study of verb inflection in a morphologically rich language (Nen)* (Unpublished Masters Thesis). Australian National University.

Nagy, N. (2009). The challenges of less commonly studied languages; writing a sociogrammar of Faetar. In J. N. Stanford & D. R. Preston (Eds.), *Variation in indigenous minority languages* (pp. 395–417). Amsterdam, The Netherlands: John Benjamins.

Nagy, N., & Meyerhoff, M. (2008). The Social Lives in Language. In M. Meyerhoff & N. Nagy (Eds.), *Social lives in language - sociolinguistics and multilingual speech communities: Celebrating the work of gillian sankoff* (pp. 1–16).

Amsterdam: John Benjamins Publishing.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. doi: 10.1111/j.2041-210x.2012.00261.x

National Statistical Office, P. N. G. (2013). *National Population and Housing Census of Papua New Guinea 2011* (Tech. Rep.). Port Moresby: National Statistical Office, Papua New Guinea.

Navarro, D. (2015). *lsr*. Retrieved from `https://www.rdocumentation.org/packages/lsr/versions/0.5`

Nettle, D. (1999). *Linguistic Diversity*. New York, USA: Oxford University Press.

Nevalainen, T., & Palander-Collin, M. (2011). Grammaticalization and sociolinguistics. In *Oxford handbook online* (pp. 119–).

Nevalainen, T., & Raumolin-Brunberg, H. (2017). *Historical Sociolinguistics: Language change in Tudor and Stuart England*. Oxon, UK: Routledge.

Nevalainen, T., Raumolin-Brunberg, H., & Mannila, H. (2011). The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change*, *23*(1), 1–43.

Newberry, M. G., Ahern, C. A., Clark, R., & Plotkin, J. B. (2017). Detecting evolutionary forces in language change. *Nature*, *551*(7679), 223–226.

Newmeyer, F. J. (1998). *Language Form and Language Function*. Cambridge, Massachusetts, USA: MIT Press.

Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago: Chicago University Press.

Nichols, J. (1995). Diachronically stable structural features. In H. Andersen (Ed.), *Historical linguistics, 1993: Selected papers from the 11th international conference on historical linguistics, los angeles, 16-20 august 1993* (pp. 337–356). Amsterdam, Netherlands: John Benjamins Publishing Company.

Nordberg, B., & Sundgren, E. (1999). *On observing real-time language change: A Swedish case study*. Uppsala: Uppsala universitet.

Nycz, J. (2015). Second Dialect Acquisition: A Sociophonetic Perspective. *Linguistics and Language Compass*, *9*(11), 469–482.

Odé, C. (2002). *Mpur prosody : an experimental-phonetic analysis with examples from two versions of the Fentora myth.* (Unpublished doctoral dissertation). Osaka Gakuin University, Osaka, Japan.

Olsson, B. (2013). *Iamitives Perfects in Southeast Asia and beyond* (Unpublished MA thesis). Stockholms Universitet.

Olsson, B. (2017). *The Coastal Marind language* (Unpublished doctoral dissertation). Nanyang Technological University.

Overall, S. E. (2007). *A Grammar of Aguaruna* (Unpublished doctoral dissertation). La Trobe University.

Paijmans, K., Blake, D. H., & Bleeker, P. (1971). Land systems ofthe Morehead-Kiunga area. In K. Paijman, D. H. Blake, P. Bleeker, & J. R. McAlpine (Eds.), *Land resources ofthe morehead-kiunga area, territory ofpapua and new guinea (land research series no. 29)* (pp. 19–45). Melbourne, Australia: Commonwealth Scientific & Industrial Research Organization, Australia.

Palmer, B. (Ed.). (2017). *The Languages and Linguistics of New Guinea: A Comprehensive Guide*. Berlin, Germany: Mouton De Gruyter.

Palmer, B. (2018). Language Families of the New Guinea Area. In B. Palmer (Ed.), *The languages and linguistics of new guinea: A comprehensive guide* (pp. 1–20). Berlin, Germany: Mouton de Gruyter.

PARADESIC. (2019). *PARADISEC (the Pacific And Regional Archive for Digital Sources in Endangered Cultures).* Retrieved from `http://www.paradisec.org.au`

Patrick, P. L. (2002). The Speech Community. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook oflanguage variation and change* (pp. 573–597). Malden, USA: Blackwell Publishing. doi: 10.1002/9780470756591.ch23

Patterson, M. (1974). Sorcery and Witchcraft in Melanesia. *Oceanic Publications*, *45*(2), 132–160.

Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Halle: Niemeyer.

Pawley, A., Gi, S. P., Majnep, I. S., & Kias, J. (2000). Hunger acts on me: the grammar and semantics of bodily and mental process expressions in Kalam. *Oceanic Linguistics*, *29*, 153–185.

Pennington, R. (2016). *A grammar of Ma Manda a Papuan language of Morobe Province, Papua New Guinea* (Unpublished PhD Thesis). James Cook University.

Perkins, R. D. (1992). *Deixis, grammar and culture*. Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Petré, P., & van de Velde, F. (2019). The real time dynamics of the individual and the community in grammaticalistion. *Language*, *94*(4), 867–900.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, *34*, 516–530.

Pierrehumbert, J. B., & Talkin, D. (1991). Lenition of /h/ and glottal stop. In *Papers in laboratory phonology ii* (pp. 90–117). Cambridge, UK: Cambridge University Press.

Podesva, R. J. (2011). Salience and the Social Meaning of Declarative Contours: Three Case Studies of Gay Professionals. *Journal of English Linguistics*, *39*(3), 233–264.

Pope, J., Meyerhoff, M., & Ladd, D. R. (2007). Forty years of language change on Martha's Vineyard. *Language*, *83*(3), 615–627.

Poplack, S. (2011). Grammaticalizacion and linguistic variation. In B. Heine & H. Narrog (Eds.), *Oxford hanbook of grammaticalization* (pp. 209–224). Oxford, UK: Oxford University Press.

Poplack, S., & Torres Cacoullos, R. (2013). Linguistic Emergence on the ground: a variationist paradigm. In M. Whinney, B. William, & O. William (Eds.), *The handbook of language emergence* (pp. 267–291). Boston, USA: Wiley-Blackwell.

Preston, D. R. (2004). *Language with an Attitude.* New Jersey, USA: Blackwell Publishing.

Rau, D. V., Chang, H.-H. A., & Dong, M.-N. (2009). A tale of two diphthongs in an indigenous minority language: Yami of Taiwan. In J. N. Stanford & D. R. Preston (Eds.), *Variation in indigenous minority languages* (pp. 259–280). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Reali, F., Chater, N., & Christiansen, M. H. (2014). The paradox of linguistic complexity and community size. In E. A. Cartmill, S. Roberts, H. Lyn, & H. Cornish (Eds.), *The evolution of language: proceedings of the 10th international conference* (pp. 270–277). Singapore.

Reesink, G. P. (1994). Domain-creating constructions in Papuan languages. In G. P. Reesink (Ed.), *Topics in descriptive papuan linguistics* (pp. 98–121). Leiden, The Netherlands: (Semaian 10). Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: how culture transformed human evolution.* Chicago, USA and London, UK: The University of Chicago Press.

Rickford, J., & Price, M. (2013). Girlz II women: Age-grading, language change and stylistic variation. *Journal of Sociolinguistics*, *17*(2), 143–179.

Roberts, S., & Winters, J. (2013). Linguistic Diversity and Traffic Accidents: Lessons from Statistical Studies of Cultural Traits. *PLoS ONE*, *8*(8). doi: 10.1371/journal.pone.0070902

Robinson, S. (2006). The Phoneme Inventory of the Aita Dialect of Rotokas. *Oceanic Linguistics*, *45*(1), 206–209.

Rodrigues, A. (2000). Panorama das línguas indígenas da Amazônia. In F. Queixalós & O. Renault-Lescur (Eds.), *As línguas amazônicas hoje* (pp. 15–28). São Paulo, Brazil: Instituto Socioambiental, Museu Parense Emílio Goeldi.

Rojas-Berscia, L. M. (2019). *From Kawapanan to Shawi: Topics in language variation and change* (Unpublishd PhD Thesis). Universiteit Nijmegen.

Romaine, S. (1999). The Grammaticalization of the Proximative in Tok Pisin. *Language*, *75*(2), 322–346.

Romaine, S. (2008). Variation in Language and Gender. *The Handbook of Language and Gender*, 98–118.

Romero, S. (2009). Phonological markedness, regional identity, and sex in Mayan: the fricativization of intervocalic /l/ in K'iche'. In J. N. Stanford & D. R. Preston (Eds.), *Variation in indigenous minority languages* (pp. 281–297). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins.

Ross, M. (1997). Social networks and kinds of speech-community event. In R. Blench & M. Spriggs (Eds.), *Archaeology and language i: Theoretical and methodological orientation* (pp. 209–261). London and New York: Routledge.

Ross, M. (2001). Contact induced change in Oceanic Languages in North-West Melanesia. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *Areal diffusion and genetic inheritance* (pp. 133–166). Oxford, UK: Oxford University Press.

Ross, M. (2005). Pronouns as a preliminary diagnostic for grouping Papuan languages. In A. Pawley, R. Attenborough, J. Golson, & R. Hide (Eds.), *Papuan pasts: cultural, linguistic and biological histories of papuan-speaking people* (pp. 15–66). Canberra: Pacific Linguistics.

Rueck, M. (2006). Social network analysis applied to language planning in the Morehead District, Papua New Guinea. *Unpublished MA thesis*.

Rumsey, A. (1993). Language and territoriality in Aboriginal Australia. In M. Walsh & C. Yallop (Eds.), *Language and culture in aboriginal australia* (pp. 191–206). Canberra, Australia: Aboriginal Studies Press.

Sahlins, M. D. (1963). Poor Man, Rich Man, Big-Man, Chief: Political Types in Melanesia and Polynesia. *Comparative Studies in Society and History*, *5*(3), 285–303.

Salisbury, R. F. (1962). Notes on Bilingualism and Linguistic Change in New Guinea. *Anthropological Linguistics*, *4*(7), 1–13.

San Roque, L., Gawne, L., Hoenigman, D., Miller, J. C., Rumsey, A., Spronck, S., ... Evans, N. (2012). Getting the story straight: Language fieldwork using a narrative problem-solving task. *Language Documentation & Conservation*, *4*, 135–174.

Sankoff, D. (1988). Sociolinguistics and Syntactic Variation. In F. J. Newmeyer (Ed.), *Linguistics: The cambridge survey* (pp. 140–161). Cambridge, UK: Cambridge University Press.

Sankoff, G. (1980a). Multilingualism in Papua New Guinea. In *The social life of language* (pp. 95–132). University of Pennsylvania Press.

Sankoff, G. (1980b). Political power and linguistic inequality in Papua New Guinea. In *The social life of language* (pp. 5–28). University of Pennsylvania Press.

Sankoff, G. (1980c). *The social life of language* (2nd ed.). USA: University of Pennsylvania Press.

Sankoff, G. (1993). Focus in Tok Pisin. In F. Byrne & D. Winford (Eds.), . Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Sankoff, G., & Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, *83*(3), 560–588. doi: 10.1353/lan.2007.0106

Sankoff, G., & Brown, P. (1976). The Origins of Syntax in Discourse: A Case Study of Tok Pisin Relatives. *Language*, *52*(3), 631–666.

Sarvasy, H. (2017). *A Grammar of Nungon: A Papuan Language of Northeast New Guinea*. Leiden, Netherlands: Brill.

Satyanath, S., & Laskar, N. B. (2009). Variable Marking in Bishnupriya Manipur. In *Variation in indigenou minority languages* (pp. 441–462). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing

Company.

Schachter, P. (1973). Focus and Relativization. *Language*, *49*(1), 19–46.

Schilling, N. (2013). *Sociolinguistic Fieldwork*. Cambridge, UK: Cambridge University Press.

Schilling-Estes, N. (2004). Constructing ethnicity in interaction. *Journal of Sociolinguistics*, *8*(2), 163–195.

Schokkin, D. (2014). Discourse practices as an areal feature in the New Guinea region ? Explorations in Paluai , an Austronesian language of the Admiralties. *Journal of Pragmatics*, *62*, 107–120. Retrieved from `http://dx.doi.org/10.1016/j.pragma.2013.09.021` doi: 10.1016/j.pragma.2013.09.021

Schokkin, D., & Lindsey, K. (2018). *Parallel variation in two speech communities: a comparison of final /n/ elision in Idi and Ende.* [Presentation given at ANU CoeDL Seminar, Canberra, Australia (August 2018)].

Seara, I. C. (2000). *Estudo acústico-perceptual da nasalidade das vogais do português brasileiro (Acoustical-perceptual study on the nasality of the vowels of Brazilian Portuguese)* (Unpublished doctoral dissertation). Universidade Federal de Santa Catarina.

Siegel, J. (2010). *Second Dialect Acquisition*. Cambridge: Cambridge University Press.

Siegel, J. (2014a). Adjectives in Nama? In *Languages of southern new guinea workshop.* Australian National University, Canbeera.

Siegel, J. (2014b). The morphology of tense and aspect in Nama, a Papuan language of southern New Guinea. *Open Linguistics*, *1*(1), 211–231.

Siegel, J. (2017). Transitive and Intransitive Verbs in Nama, a Papuan language of Southern New Guinea. *Oceanic Linguistics*, *56*(1).

Siegel, J. (accepted). The Relative Pronoun strategy revisited: New data from southern New Guinea. *Studies in Language*.

Sillitoe, P. (1998). *An introduction to the anthropology of Melanesia: culture and tradition*. Cambridge, UK: Cambridge University Press.

Silverstein, M. (2015). How Language Communities Intersect : Is"superdiversity" an incremental or transformative condition? *Language and Communication*, *44*, 7–18.

Simons, G. F. (1982). Word Taboo and Comparative Austronesian Linguistics. In A. Halim, L. Carrington, & S. A. Wurm (Eds.), *Papers from the third international conference on austronesian linguistics* (pp. 157–226). Pacific

Linguistics, Series C, 76, Vol.3.

Simpson, A. P., & Ericsdotter, C. (2003). Sex-Specific Durational Differences in English and Swedish. *15th ICphs*, 1113–1116.

Simpson, G. G. (1944). *Tempo and mode in evolution.* New York, USA: Colombia University Press.

Singer, R. (2018). A small speech community with many small languages: The role of receptive multilingualism in supporting linguistic diversity at Warruwi Community (Australia). *Language and Communication*, *62*, 102–118.

Singer, R., & Harris, S. (2016). What practices and ideologies support small-scale multilingualism? A case study of Warruwi Community, northern Australia. *International Journal of the Sociology of Language*, *2016*(241), 163–208. doi: 10.1515/ijsl-2016-0029

Sinnemäki, K. (2009). Complexity in core argument marking and population size. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 126–140). Oxford University Press.

Smeall, C. (1975). Quantitative Analysis of Variation : i in Tok Pisin. *Proceedings of the First Annual Meeting of the Berkeley Linguistics*, 403–409.

Smith, A. D. M., Trousdale, G., & Waltereit, R. (2015). Introduction. In A. D. M. Smith, G. Trousdale, & R. Waltereit (Eds.), *New directions in grammaticalization research.* Amsterdam, Netherlands: John Benjamins Publishing Company.

Sober, E. (1980). Population Thinking , and Essentialism. *Philosophy of Science*, *47*(3), 350–383.

Solheim, R. (2009). Dialect development in a melting pot: The formation of a new culture and a new dialect in the industrial town of Høyanger. *Nordic Journal of Linguistics*, *32*(2), 191–206.

Stanford, J. N. (2007). *Dialect contact and identity: a case study of exogamous Sui clans* (Unpublished PhD Thesis). Michigan State University.

Stanford, J. N. (2008). A sociotonetic analysis of Sui dialect contact. *Language Variation and Change*, *20*(3), 409–450. doi: 10.1017/S0954394508000161

Stanford, J. N. (2009a). Clan as a sociolinguistic variable: Three approaches to Sui clans. In J. N. Stanford & D. R. Preston (Eds.), *Variation in indigenous minority languages* (pp. 463–484). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins.

Stanford, J. N. (2009b). "Eating the food of our place": Sociolinguistic loyalties in multidialectal Sui villages. *Language in Society*, *38*(03), 287.

Stanford, J. N., & Pan, Y. (2013). The sociolinguistics of exogamy: Dialect acquisition in a Zhuang village. *Journal of Sociolinguistics*, *17*(5), 573–607. doi: 10.1111/josl.12052

Stanford, J. N., & Preston, D. R. (2009). The Lure of a Distant Horizon: Variation in indigenous minority languages. In J. N. Stanford & D. R. Preston (Eds.), *Variation in indigenous minority languages* (pp. 1–20). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Stanford, J. N., Severance, N. A., & Baclawski, K. P. (2014). Multiple vectors of unidirectional dialect change in eastern New England. *Language Variation and Change*, *26*(1), 103–140. doi: 10.1017/S0954394513000227

Stasch, R. (2009). *Society of Others: Kinship and mourning in a West Papuan Place*. Berkeley, USA: University of California Press.

Stasch, R. (2011). Word Avoidance as a Relation-Making Act: A Paradigm for Analysis of Name Utterance Taboos. *Anthropological Quarterly*, *84*(1), 101–120.

Steed, W., & Hardie, P. (2004). Acoustic Properties of the Kuman Voiceless Velar Lateral Fricative. *Proceedings of the 10th Australasian International Conference on Speech Science and Technology. Sydney: Australian Speech Science and Technology Association*, 346–351.

Stirling, L. (2012). Tense/aspect shifting in Kala Lagaw Ya. *Australian Journal of Linguistics*, *32*(1), 157–190.

Summerhayes, G. R., Leavesley, M., Fairbairn, A., Mandui, H., Field, J., Ford, A., & Fullagar, R. (2010). Human adaptation and plant use in highland New Guinea 49,000 to 44,000 years ago. *Science*, *330*(6000), 78–81.

Sutton, P. J. (1978). *Wik: Aboriginal society, territory and language at Cape Keerweer, Cape York Penninsula, Australia* (Unpublished doctoral dissertation).

Sutton, P. J. (1997). Materialism, sacred myth and pluralism: competing theories of the origin of Australian Languages. In F. Merlan, A. Rumsey, & J. Morton (Eds.), *Scholar and sceptic: Australian aboriginal studies in honour of l. r. hiatt* (pp. 21–42). Canberra, Australia: Aboriginal Studies Press.

Tabain, M., Fletcher, J., & Butcher, A. (2014). Lexical stress in Pitjantjatjara. *Journal of Phonetics*, *42*(1), 52–66.

Tadmor, U., Haspelmath, M., & Taylor, B. (2010). Borrowability and the notion

of basic vocabulary. *Diachronica*, *27*(2), 226–246. doi: 10.1075/dia.27.2.04tad

Tagliamonte, S. A. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation.* West Sussex, United Kingdom: Wiley-Blackwell.

Tagliamonte, S. A., & Denis, D. (2008). Linguistic ruin? Lol! Instant messaging and teen language. *American Speech*, *83*(1), 3–34.

Team, R. C. (2017). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from `https://www.r-project.org/`

ten Thije, J. D., & Zeevaert, L. (2007). *Receptive Multilingualism : Linguistic analyses, language policies and didactic concepts.* Philadelphia: John Benjamins Publishing Co.

Thomas, E. R. (2015). Sociophonetics of Consonantal Variation. *Annual Review of Linguistics*, *2*(1), 95–113.

Thomason, S. G. (2007). Language Contact and Deliberate Change. *Journal of Language Contact*, *1*(1), 41–62. doi: 10.1163/000000007792548387

Thompson, S. A., Longacre, R. E., & Hwang, S. J. (2007). Adverbial Clauses. In T. Shopen (Ed.), *Language typology and syntactic description second edition: Volume 2: Complex constructions* (2nd ed., pp. 237–300). Cambridge, UK: Cambridge University Press.

Thurston, W. R. (1987). *Processes of Change in the Languages of North-Western New Britain* (Vol. B-99). Canberra, Australia: Pacific Linguistics.

Thurston, W. R. (1992). *Sociolinguistic typology and other factors effecting change in north-western New Britain, Papua New Guinea* (No. C). Canberra: Pacific Linguistics.

Torres Cacoullos, R., & Walker, J. A. (2009). On the persistence of grammar in discourse formulas: A variationist study of that. *Linguistics*, *47*(1), 1–43.

Traugott, E. C. (1989). On the Rise of Epistemic Meanings in English : An Example of Subjectification in Semantic Change. *Language*, *65*(1), 31–55.

Traugott, E. C., & Dasher, R. B. (2002). *Regularity in Semantic Change.* Cambridge, UK: Cambridge University Press.

Treis, Y. (2005). Avoiding Their Names , Avoiding Their Eyes : How Kambaata Women Respect Their In-Laws. *The Trustees of Indiana University on behalf of Anthropological Linguistics*, *47*(3), 292–320.

Trudgill, P. (1974). *The Social Differentiation of English in Norwich.* Cambridge, UK: Cambridge University Press.

Trudgill, P. (1983). *On Dialect*. Oxford, UK: Blackwell.

Trudgill, P. (2004a). *New-dialect formation: The inevitability of colonial Englishes*. Edinburgh, UK: Edinburgh University Press.

Trudgill, P. (2004b). *New-dialect Formation: The Inevitability of Colonial Englishes*. Edinburgh, UK: Edinburgh University Press.

Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford, UK: Oxford University Press.

Tryon, D. T. (1976). New Hebrides Languages: An Internal Classification. *Pacific Linguistics*, *C-50*.

Tucker, S., Boevé, M., Fuller, E., Gustafsson, C., & Rueck, M. (2003). *Nambu Subfamily survey report* (Tech. Rep.). Ukarumpa: Summer Institute of Linguistics.

Urame, J. (2015). The Spread of Sorcery Killing and its Social Implications. In M. Forsythe & R. Eevs (Eds.), *Talking it through: Responses to sorcery and witchcraft beliefs and practices in melanesia* (pp. 23–36). Canberra, Australia: ANU Press.

van. Baal, J. (1966). *Dema, description and analysis of Marind-Anim culture (South New Guinea)*. The Hague, Netherlands: Martinus Nijhofd.

van Hofwegen, J. (2011). Apparent-time evolution of / l / in one African American community. *Language Variation and Change*, *22*(3), 373–396.

Vaughan, J. (2018). "We talk in saltwater words": Dimensionalisation of dialectal variation in multilingual Arnhem Land. *Language and Communication*, *62*, 119–132.

Vaughan, J., & Singer, R. (2018). Indigenous multilingualisms past and present. *Language and Communication*, *62*, 83–90.

Verschik, A. (2012). Practising receptive multilingualism: Estonian-Finnish communication in Tallinn. *International Journal of Bilingualism*, *16*(3), 265–286. doi: 10.1177/1367006911426465

von Mengden, F. (2010). *Cardinal Numerals : Old English from a Cross-Linguistic Perspective*. De Gruyter.

Wagner, R. (1984). Rituals as communication: order, meaning, and secrecy in Melanesian initiation rites. *Annual Review of Anthropology*, *12*, 143–155.

Wagner, S. E., & Sankoff, G. (2011). Age grading in the Montréal French inflected future. *Language Variation and Change*, *23*(03), 275–313. doi: 10.1017/s0954394511000111

Walker, J. A. (2012). Form, function, and frequency in phonological variation.

Language Variation and Change, *24*(2012), 397–415.

Walker, J. A., & Meyerhoff, M. (2013). Studies of the Community and the Individual. In R. Bayley, R. Cameron, & C. Lucas (Eds.), *Oxford handbook of sociolinguistics* (pp. 175–194). doi: 10.1093/oxfordhb/9780199744084.013.0009

Ward, N. (2006). Non-lexical conversational sounds in American English. *Pragmatics & Cognition*, *14*(1), 129 –182.

Warren, P. (1998). Timing patterns in New Zealand English rhythm. *Te Reo*, *41*, 80–93.

Wassink, A. B. (1999). *A Sociophonetic Analysis of Jamaican Vowels* (Unpublished PhD Thesis). University of Michigan.

Wassink, A. B. (2006). A geometric representation of spectral and temporal vowel features: quantification of vowel overlap in three linguistic varieties. *The Journal of the Acoustical Society of America*, *119*(4), 2334–2350.

Weinreich, U., Labov, W., & Herzog, M. (1968). *Empirical Foundations for a Theory of Language Change*. Texas, USA: University of Texas Press.

Wells, J. C. (1982). *Accents of English*. Cambridge, UK: Cambridge University Press.

Whiteside, S. P. (1996). Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. *Journal of the International Phonetic Association*, *26*(1), 23–40.

Wichmann, S., & Holman, E. W. (2009). *Temporal stability of linguistic typological features*. München: LINCOM Europa.

Williams, D., & Escudero, P. (2014, nov). A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the Acoustical Society of America*, *136*(5), 2751–61.

Williams, F. E. (1936). *Papuans of the Trans-Fly*. Oxford: Clarendon Press.

Wilson, J. (1994). Paradoxes, sociolinguistics and everyday accounts. *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, *13*(3), 285–300.

Wilson, J. (2017). *A Grammar of Yeri a Torricelli language of Papua New Guinea by* (Unpublished PhD Thesis). State University of New York.

Wilson, P. (1980). *Ambulas grammar* (Vol. 26). Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Wolfram, W. (1969). *A sociolinguistic description of Detroit Negro speech*. Washington DC: Center for Applied Linguistics.

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, *117*(3), 543–578.

Wurm, S. A. (1975). *New Guinea area languages and language study* (Vol. 1). Canberra, Australia: Dept. of Linguistics, Research School of Pacific Studies, Australian National University.

Yaegr-Dror, M. (1996). Phonetic evidence for the evolution of lexical classes: the case of Montreal French vowel shift. In G. R. Guy, C. Feagin, D. Schiffrin, & J. Baugh (Eds.), *Towards a social science of language: Papers in honor of william labov, volume 1* (pp. 263–288). Amsterdam, the Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Yang, C., Stanford, J. N., & Yang, L. (2019). Variation in the tonal space of Yangliu Lalo, an endangered language of Yunnan, China. *Linguistics of the Tibeto-Burman Area*, *42*(1), 2–37.

Zhang, Q. (2005). A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in Society*, *34*(3), 431–466.

Zwicky, A. M., & Pullum, G. K. (1983). Cliticization vs. Inflection: English N'T. *Language*, *59*(3), 502–513.